Optimization-Based Approach
to Classification Tasks

Mat-2.108 Sovelletun matematiikan erikoistyö

Ilkka Männistö Tf 48474F
# Table of contents

1 Introduction .......................................................................................................................... 4

2 Types of Classification Tasks ............................................................................................ 5
   2.1 Overview .......................................................................................................................... 5
   2.2 Generalization and Pattern Recognition ...................................................................... 6
   2.3 Organization of Information ......................................................................................... 6
   2.4 Ranking Systems .......................................................................................................... 7
   2.5 Resource Allocation ..................................................................................................... 7
      2.5.1 Resource allocation under risk ........................................................................... 9
   2.5.2 Burden-to-importance ratio (BIR) .......................................................................... 10
   2.5.3 Item Life Span and Reclassification ...................................................................... 10
   2.6 Summary of Different Classification Tasks ................................................................. 11

3 Examples of Classification Problems ............................................................................... 13
   3.1 Prisoner Security Classification .................................................................................. 13
      3.1.1 Another Kind of Prisoner’s Dilemma................................................................. 13
      3.1.2 Inmate Custody Levels and Prison Security Levels ........................................... 13
      3.1.3 Optimal Classification ....................................................................................... 14
   3.2 Document Security Classification ............................................................................... 15
      3.2.1 Overview .............................................................................................................. 15
      3.2.2 Security Classes ................................................................................................... 15
      3.2.3 Classification Process ......................................................................................... 16
      3.2.4 Downgrading and Declassifying Information ...................................................... 17
   3.3 EBRCM – Experience Based Reliability Centered Maintenance ................................ 17
      3.3.1 Overview .............................................................................................................. 17
      3.3.2 Maintenance Strategy Classification ..................................................................... 18

4 Optimal Classification Example ....................................................................................... 19
   4.1 Basics ........................................................................................................................... 19
   4.2 Constructing the Model ............................................................................................... 19
   4.3 Optimization ................................................................................................................. 21
      4.3.1 Multi-objective Optimization .............................................................................. 21
      4.3.2 Calculating the Pareto-optimal Curve ................................................................. 22
      4.3.3 Classifying New Components ............................................................................. 22
      4.3.4 The Role of the Middle Alternative ...................................................................... 24
      4.3.5 Combining Optimization Objectives ................................................................... 26

5 Discussion and Conclusions ............................................................................................. 27

Appendix A ............................................................................................................................. 28
   Proof of BIR-measure ........................................................................................................... 28
Appendix B

Constructing the Pareto-optimal Set

Appendix C

Proof for Viability of Middle Class
1 Introduction

In context of this study classification tasks are considered to be such mathematical mappings where the items to be classified are each assigned to certain class. In practice classification tasks are such that there is a large number of items compared to the number of possible classes. The decision mechanic for choosing the right class depends on the system, as will be demonstrated in this study. This approach is similar to a basic view of pattern recognition, in [3]. A model for more general classification problem is presented in this study, one that is broader than the standard pattern recognition problem.

Classification tasks have been used to understand, manage and enhance a wide variety of real world processes and systems. These applications can range from assessing ships seaworthiness to organizing libraries. These are widely varying systems, except for one thing: they all can be very large and complex systems, given the large number of different kinds of ships and books in existence. Both of these systems have been simplified by classifying the ships or books into broader groups. Books are placed in different sections in library, and ships have classes for insurance purposes, for example. Finding anything in library would be more difficult without sections, as would be pricing insurance for a vessel if those were handled individually.

Classification can also be used for a greater understanding of a complex system. Qualities and relations inherent to the system being analyzed are many times more apparent when a broader class is examined, rather than individual items. Understanding a system is of course essential in the process of making it more efficient. Classification is not a purpose in itself – just a tool in improving or making viable another process.

While giving an overview of the different types of classifications the main focus of this study is on the more mathematically viable classification tasks. This means that the goals of applying the classification scheme to the process must be well quantized, but this also allows us to efficiently use other mathematical tools. Optimization and decision theory are the two main tools used in this study.

This study consists of an overview of different classifications tasks ranging from pattern recognition to resource allocation, presented in section 2. A few specific examples of tried and tested classification systems are examined in section 3 in detail. One of those examples is mathematically analyzed in section 4. Conclusions and discussion in section 5 wraps up the study.
2 Types of Classification Tasks

2.1 Overview

The basic classification problem is to assign N items in terms of K classes based on the attributes of the item in question. This approach is also taken in main works in pattern recognition [1] [2]. Classification problems in this study are of a broader form than straight pattern recognition problems. In pattern recognition there is always a right class for the item to belong into. As is shown in this classification tasks, that is not always the case.

The attributes can be measured numerical features of the item, like in pattern recognition, or they can be determined by an expert panel or they can even be just qualitative data. In some cases the most complicated part in classification can be the extraction of attributes from the information on the item to be classified. Multivariate analysis can be used as a classification method in cases of uncertain observations or measurements of attributes [4].

The following picture gives an outline of the classification process:

![Classification process diagram]

**Figure 1: Classification process, modified from [2]**

This model of classification process can be used to illustrate widely varying classification tasks. The box with text “Model of the System” illustrates that the underlying mathematical
model of the system determines and affects each part of the process, as noted with the two-way arrows.

In the following sections classification tasks are broken down to different categories by their qualities.

2.2 Generalization and Pattern Recognition

Certain types of pattern recognition problems can be seen as classification tasks. For example recognizing letters in a written language is to assign each shape to the, hopefully, correct class which is in this case the right letter. All the different styles and shapes of the letter ‘a’ are to be classified as the letter a, and so forth. Without this generalization it would not be possible to construct the grammar for any language – all the different variations (due to different handwritings, different fonts, etc) of letter ‘a’ would have to be treated as different items. Also in grammar words are classified in to word classes too, so that the grammar itself can be constructed. Generalization thus allows us to make rules and teach others.

These classification tasks also form a part of human learning. Early humans might have tried different kinds of berries, and find out that others nauseate and others are delicious. Humans then start to recognize berries, and before long will know before eating if they are nauseating or not. This too is a classification process, in which the berries are classified into edible and non-edible ones. The attributes of the berries are for example color, shape and size. It must be noted, however, that human learning process is vastly more complicated than simply classifying items in to different classes [5]. Classification of information is only a part of the cognitive process in learning.

2.3 Organization of Information

In organization of information classification of items is used to manage large number of items. The classes are not ranked, but are ordered to make the information more readily accessible. That means to say that the different classes do not have any special relations between each other. Libraries organize books by their subject – cooking books are grouped in a different section than history books, and history books can further be divided to European and American history. The classification task here is to decide in which category certain book belongs, judged by its content. Dewey Decimal is one such classification system [6]. This type of classification is also used in internet directories, like Yahoo.

Classification can also be used to help finding the underlying principles or laws in different subjects. For example, taxonomy is the science of classifying and determining the evolution of life forms. With the right classification life in different branches in the “tree of life” it is possible to construct a chain of evolution from ancestral organisms to modern day animals. Taxonomy is also an example of another hierarchical process – life is classified first in broader categories, like vertabrae, and then into narrower classes until the specific race is found. In this case the attributes of the single life form are certain physiological features it possesses, like number of legs, body configuration and composition [7].

The process can also be more freeform. In all cases there necessarily are not any groups given to which the items should be classified – rather, it is the purpose of the classification to find
the different groups and differentiate between them. Clustering is one such method. The purpose of cluster analysis is to place objects into groups or clusters suggested by the data, not defined a priori, so that objects in the same cluster tend to be similar and objects in different clusters dissimilar [8]. This can give insight to the data or at least summarize it. A marketer could for example recognize a group of consumers in his clientele that has not been recognized before. This would allow him to further exploit the clientele by more suitable marketing. There are several different types of clustering techniques: disjoint clusters, hierarchical clusters, overlapping clusters and fuzzy clusters. These different types of clusters are defined by how the groups can overlap. There are also several different algorithms for clustering.

2.4 Ranking Systems

In the classification tasks examined previously the classes did not have an order or rank that would allow us to compare different classes, i.e., they are nominal. In some cases the class also is a rank or a measure of certain characteristic of the items in the class. Thus, the classes can be arranged according to the class. This ranking can be either qualitative or quantitative and absolute or relative. Even though the goal of the classification task might be optimizing the risks involved or use of resources, one inherent advantage of using classification is that it makes the system manageable in the first place. In many cases there are far too many items involved to give each an individual treatment, and even if it was, it would be hard to get a comprehensive overview of the system.

Ranking of items is a classification task where the relevant features of the item are combined into a single measure. Measures are then distributed to K discrete ranks. Grading of exams is such a problem. There might be only 5 possible grades, but the exams done by individual students have a much finer scale, like points gained. Car crash tests can also be seen as similar processes. 5 stars, 4 stars, etc, can be thought as different classes, and the results from different impact tests are the attributes of the item in question. Classification to different classes then depends on the guidelines on what the items in each class must accomplish.

Ranking can be a tool for resource allocation problems, which are further discussed in section 2.5. The items rank can be based on its importance to the system – and the item's importance is decisive resource allocation. This is useful when there is a very large number of items to be classified. Ranking is an intermediate phase before the final allocation.

2.5 Resource Allocation

In cases where a cost (or a use of limited resources) is attached to each class a possible goal of that classification task is an optimal distribution of the items to different classes. Many real world classification problems are of this type. In designing a classification system for resource allocation first a mathematical model must be constructed to link together the item's attributes and its importance. Item's importance to the system is deduced from its attributes, and then it is assigned to the optimal class based on the overall cost function. The importance depends on the goals of the classification task and the system to which the item to be classified belongs.
For example the importance can be the risk an inmate poses to the prison, or the importance of certain component in a power plant or just a general rank decided by a panel of experts. Experience based reliability centered maintenance works in that manner: expert panel determines the importance of certain part of equipment to the power plant from information available, and then rank it accordingly [12]. The rank helps then to assign correct maintenance procedures to the equipment in question.

The basic assumption is that we have N items we manage, and K different procedures (different maintenance actions, etc.) which we can apply to each item. Applying a procedure to an item results in a gain or loss that depends on both the item and procedure, and a drain on resources that depends on the procedure done. For example in a hospital its resource could be its available operation room time, from which each different medical procedure uses a certain time independent of to whom the operation is done. The gain or loss is the patients' wellbeing, which depends on both the procedure and the patient. While this problem could be formulated as a classification task, it might be difficult to define and operate wellbeing in monetary terms because of ethical considerations. Prison fits the same formula as well. Prisoners are the items in this case, procedures are the different security level wards and gain / loss is the risk the prisoner causes in each security level.

Formally we can present the resource allocation problem as follows:

\[ V(\bar{x}, \bar{c}) = W_1 \sum_{i} C(c_i) + W_2 R(\bar{x}, \bar{c}) \]  
(2.1)

with respect to \( \bar{c} = (c_1, ..., c_N) \), \( c_i \in \{1, ..., K\} \)

Subject to constraints:
\[ \sum C(c_i) < C_{max} \]  
(2.2)
\[ R(\bar{x}, \bar{c}) < R_{max} \]  
(2.3)

Where
\[ C(c) = \text{The drain on resources caused by a single item when classified into class d. Assumed to be independent of item to be classified.} \]

\[ R(\bar{x}, \bar{c}) = \text{Function portraying the overall risk of the system.} \]

\[ W_1, W_2 = \text{Decisionmakers’ weights for resources (W_1}) \text{ and risk (W_2).} \]

\[ \bar{c} = (c_1, ..., c_N) \text{ vector in which } d_j \text{ is the class item } j \text{ is assigned to.} \]

\[ \bar{x} = (x_1, ..., x_N) \text{ information on items.} \]

### 2.5.1 Resource allocation under risk

Classification systems with the goal of risk management or minimization usually belong to this category. Items classified in such systems present a certain amount of risk, depending on which measures are taken to control that risk. All items in a certain class have similar measures to control the risk. The classification problem is then to assign each item to a class in a way that minimizes the overall risk. In this case the advantages of classification are twofold: risk is minimized and it can be managed in a comprehensive way. Security classes are an example of this, and will be further examined in the next section.

In risk minimization the utility function \( f(x,c) \) is a probability distribution that portrays the possible losses item \( x \) might cause when assigned to class \( c \). To construct \( f(x,c) \) we need to analyze what negative outcomes there are in the system, the probability of each outcome, their cost to the system and how these variables depend on the items attributes and the class it is assigned into. In formal terms the probability is a function \( p_i(x,c) \) and the cost associated with is \( u(D_i,c) \), where \( i \) is the index of the negative outcome, \( D_i \) is the cost of the negative outcome \( i \) and \( u \) is a utility function. Thus \( f(x,c) \) can then be written as:

\[
 f(x,c) = \sum_{i=1}^{n} p_i(x,c)u(D_i,c) \tag{2.4}
\]

if there are \( n \) different, known, negative outcomes in the system. This presents the expected utility of that particular classification. Different estimates than expected value could also be used. The whole risk minimization problem can be written as a classification problem as:

\[
 \min \sum_{j=1}^{N} \sum_{i=1}^{n} p(x_,c)u(D_i,c) \tag{2.5}
\]

subject to:

\[
 \sum_{k=1}^{N} s(c_k) \leq R \tag{2.6}
\]

where \( R \) is the total amount of resources available and \( s(c_k) \) is the amount of resources expended when classifying a single item into class \( c \). It is also possible that the amount of resources depended on the item in question, so that \( s(c) \) is actually \( s(c,x) \).
Equation (2.5) is only valid in a risk-neutral situation, where the smaller risks from different item and class pairs combination can be added together, i.e., the utility of the system is the sum of the utilities of individual items. To take into account different risk attitudes a utility function would have to be added to the equation.

2.5.2 Burden-to-importance ratio (BIR)

One method for evaluating the effectiveness of the resource allocation is the burden-to-importance ratio. In (2.5) we calculated the overall risk in the system. Each item classified in the system contributes to this risk according to (2.4). The ratio of these two figures is the relative importance of the item. Function $s(c_i)$ tells us the drain on resources, or burden, item $i$ causes in class $c$. The total burden all items cause in the system is equal to equation (2.2) or, usually $R$ if all resources are used. Relative burden of the item is the ratio of total burden and the individual burden the item causes. Burden to importance ratio then is the relative burden divided by relative importance:

$$BIR = \frac{RB}{RI}$$

When the resources are allocated effectively, burden to importance ratio (BIR) should be near 1. This means that the resources expended are equal to the risk importance of the item in question. Resources can mean any limited assets that are expended in reducing the risk, like money, man-hours or spare parts that are in this classification scheme dependent on the class to which the item is classified. Maximizing the effectiveness of resources automatically leads to BIR of 1 [9]. When using BIR principle in classifying system, where the resources expended depend on the classes, it is not possible to achieve a BIR of 1 for each item. This is because there are only discrete number of alternatives. The proof of BIR value being 1 in optimal resource allocation is in Appendix A.

2.5.3 Item Life Span and Reclassification

Items are classified to a certain class based on their attributes upon classification. These attributes may change over time, and it may be efficient to change the classification later on. For example in maintenance it is important to consider equipment life span when assigning maintenance classes. A component nearing the end of its life span may not be worth the resources it’ll consume in the highest maintenance class. Gathering information on the classified items also costs or uses resources, so it is likely that the whole system cannot be constantly monitored and evaluated; rather, any reclassifications are done at certain intervals. Length of those intervals is another decision problem, and depends on the gains of reclassification.

In reliability analysis the different phases of component lifetime have been identified as burn-in, useful life and wear-out periods. Burn-in phase is the time directly after the new machine has been taken into use. Failure rate is often high in this period because of hidden defects, which show up quickly after the unit has been taken in to use. When these failures have been identified and fixed the failure rate often stabilises to a new, lower level. This is the useful life period, which lasts until the component begins to wear out. As the component nears the end of
its life the failure rate starts to rise again. Most component failure rates follow this so-called bathtub curve, which is shown in picture 2.1 [13], where $z(t)$ is the failure rate as a function of time.

![Bathtub Curve Diagram](image)

### Figure 2.1 Bathtub curve [13]

In a classification scheme in which classes determine the maintenance action, optimal classification might call for reclassification at different times in the items' life span. For example, in the above picture a higher maintenance class might be warranted in first and third periods. This follows from the heightened risk importance due to higher failure rate. Efficient maintenance requires that more resources be spent on more important items, as was discussed in section 2.5.1 burden-to-importance ratio.

System modifications can also give the need for reclassification. When the system is modified the relative importances of the classified items may change, and the current classification might not be optimal.

Another reason for a dynamic change in the process is that the classes may change. Costs associated with certain class may change, or the meaning of the class may change. This will likely lead to re-evaluation of the whole classification process, since possibly all items will be affected, unlike in the case of changed attributes in a single item, above. For example a price jump in certain types of spare parts could bring up the drain of resources in certain maintenance classes while others might be unaffected. This could likely lead to a new optimal classification.

### 2.6 Summary of Different Classification Tasks

The different classification tasks presented in this section can be differentiated in regards to their main characteristics. The following table has a summary of all the different tasks:
<table>
<thead>
<tr>
<th>Classification Task</th>
<th>Purpose of classification</th>
<th>Classification Criterion</th>
<th>Constraints</th>
<th>Classes ordered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pattern Recognition</td>
<td>Find right solution/not to make mistake</td>
<td>Error function</td>
<td>Error limits</td>
<td>No</td>
</tr>
<tr>
<td>Organization of Information</td>
<td>Minimization of search effort</td>
<td>Similarity criterion</td>
<td>Inherent to task/decided</td>
<td>No</td>
</tr>
<tr>
<td>Resource Allocation</td>
<td>Optimal resource use</td>
<td>Utility function</td>
<td>Resource and utility function limits</td>
<td>Yes</td>
</tr>
<tr>
<td>Ranking Systems</td>
<td>Optimal differentiation</td>
<td>Utility function</td>
<td>Decided</td>
<td>Yes</td>
</tr>
</tbody>
</table>
3 Examples of Classification Problems

3.1 Prisoner Security Classification

3.1.1 Another Kind of Prisoner’s Dilemma

*Prisoner’s Dilemma* is a well-known problem in game theory, used to demonstrate Nash equilibrium in a simple case of two prisoners pondering whether to give up each other. In today’s world another kind of problem involving prisoners is found in the United States: managing the vast number of prison inmates. In 2001 over 2 million people (up 0.4 million from 1996) were serving time as inmates in prisons or jails, and a total of 6.6 million people (up 1.0 million from 1996) [11] were to some degree in the care of US Department of Justice. Growing prisoner population combined with limited resources has lead to searching for increasingly efficient inmate management models in correctional facilities.

Inmates, like all people, are all different, with different attitudes, personalities and convictions. As each individual inmate cannot realistically be given individual treatment, it is also unreasonable to treat each inmate in the same way. Each prisoner poses a different degree of risk to the prison staff, to other inmates and a different risk of flight. In light of these things one might consider it useful to somehow classify prisoners in different inmate custody levels to achieve optimal use of the limited resources while maintaining minimum risk of disturbances in the prisons. Different custody levels would then have different security measures – more guards per prisoner, limited interaction and activities with other prisoners, more secure buildings and the like.

3.1.2 Inmate Custody Levels and Prison Security Levels

Classifying prisoners is not a new idea as such (maximum security prisons have been around for ages, implying that some prisoners were considered more dangerous than others), but it is only in the 1990’s that a drive for more professional and efficient working practices allowed for real improvement in the field. In designing any classification scheme one must first assert what goals are to be attained with the system, what information is available and what the available resources are.

To increase safety to inmates and staff, more dangerous inmates should be subjected to more strict security measures. The classification process must then take into account the danger each inmate is thought to present to the system, and then assign that criminal to a suitable inmate custody level. The inmate custody rating has 4-6 different levels; for example in North Carolina there are levels close, medium, minimum I, minimum II and minimum III.

Prisons are classified by their security level, which can be for example close, medium or minimum. Maximum security is a rating even more secure than close, and is applied to certain cellblocks within close security prisons. Security levels are determined by the design and unique features of the prison, the level of staffing, and the operating procedures. Maximum security is the most restrictive level of confinement and minimum security is the least...
restrictive. The prison security level is an indicator of the extent to which an offender who is assigned to that facility is separated from the civilian community.

### 3.1.3 Optimal Classification

Three things must be considered when formalizing the inmate classification problem: goals, resources and information. Goals tell us what we actually trying to achieve with the classification. The goals for an inmate classification system could be viewed as the following:

- Safety of inmates and staff
- Public safety
- Protection against liability and inmates rights

Second, we need data on the inmates, and a mathematical risk model on how that data affects the goals set earlier. Common sense tells that criminals with convictions for violent crimes will pose a greater risk to staff and other inmates, for example. Previously inmate classification depended on just that; the common sense of the prison officials. Today much more information is utilized in classifying prisoners. They are put through a series of evaluations, including medical and mental health screenings. Prison classification specialists develop an individual profile of each inmate that includes the offender’s crime, social background, education, job skills and work history, health, and criminal record, including prior prison sentences. Based on this information, the offender is assigned to the most appropriate custody classification [11].

Prisons have their own security ratings, and the resulting overall risk the inmate causes is a function of that rating and the prisoners own custody level. Inmates have different sensitivity to the prison security level – the risk caused by a dangerous inmate varies depends greatly on the security level of the prison he is in, but a very benevolent prisoner will cause very little risk in a prison of any security level. Like other resource allocation problems this too comes down to burden-to-importance ratio. Resources expended on the prisoner should be proportional to the importance, or in this case the risk or dangerousness, of the prisoner.

As the objective of the classification system is to fulfill the goals above, the classification system should minimize overall risk. Constraints to the optimization problem are the prisons’ security ratings and their respective capacities. Monetary costs can also be added to the equations – both as constraint on the resources used, and as second optimization objective.

Prisoner classification is also a dynamic process. Inmates have a certain term in the prison, some leave the prison and some do not necessarily need to be in the same kind of security for the whole time. Therefore prisoners are evaluated at certain intervals during their prison terms. Good habits, nonviolent behaviour and such can earn a reduction in custody level (more privileges) and eventually a transfer to a prison with lesser security rating. Negative activities will of course bring opposite results.
3.2 Document Security Classification

3.2.1 Overview

Every government in the world is responsible for maintaining national security. A part of national security is to keep some important or harmful information from general knowledge, or in other words, *secret*. Such information might be information about troop movements, new weapons technology, political activities, intelligence activities or cryptology. In the case of new weapons technology, for example, it is useful to keep information on the weapon secret, because that will prevent the enemy from developing counter-measures against it, and will ensure a surprise use of the new weapon. In political negotiations it is in some cases helpful to keep some of the objectives or motives secret. From just these examples we know that there are some cases where making information secret is useful. However, keeping secret information that does not need to be secret is not good either because security is expensive. The logical question then is how much and how secret should we make information?

Classifying documents to certain groups that have similar needs for security is the commonly used method for managing national security. The following is based on the U.S classification system for nuclear technology [10].

3.2.2 Security Classes

There are only two ways to keep a secret *perfectly* safe: either tell it to no one, or use the Captain Kidd method. These methods fall outside the classification system. Normally people have accepted that perfect security is not attainable, but that there are different degrees of security. As individual treatment of each document is not possible, there are only a few distinct classes for security, namely confidential, secret and top secret. Confidential is the least secure and top secret is the most secure class. Each of the security classes has different regulations for handling the material.

The risk governments want to protect against is unwanted disclosure. This usually means that the information is made either public knowledge or that the information falls into the enemies hands. Probability of unwanted disclosure is greater in the classes that have lesser security procedures, but generally this also means that its cheaper to process documents with lower security classes. Hand in hand with the security classes go security ratings. Security ratings are permits for people to access the classified information. This means that people without the appropriate rating cannot access the classified information. Probability of unwanted disclosure is linearly proportional to the number of people that have access to the information:

\[
PDD = 1 - (1 - k_1)^{NP} \approx 1 - (1 - k_1 \times NP) = k_1 \times NP
\]

*PDD* is the probability of deliberate unauthorized disclosure. *NP* is the number of people with access to the information. It is assumed that each person has the same probability of divulging the information (if this was not true, and the chances differed, the person would not be given a security rating), which is here represented by a disloyalty factor *k*, which is in the range of $10^{-5}$ (one spy per 100,000 people). In higher security classes the probabilities are

---

* Captain Kidd was a pirate who was said to bury his treasure for safekeeping and later recovery. He is alleged to have usually killed the persons who helped him bury this treasure to keep them from revealing its location to others or from returning on their own to dig it up.
smaller: fewer and more loyal people have higher security ratings. In above equation that means lower values for each $k_i$ and $NP$, resulting in lesser probability.

3.2.3 Classification Process

Risks of information disclosure are not as clear or as quantizeable as in the inmates’ case. The main reason for classification is that the information at hand could damage national security. National security itself is a loose term, and it ‘includes all matters that directly or may reasonably be connected with the defense of the United States against any of its enemies’. It refers to the military and naval establishments and the related activities of national preparedness.

The whole classification process is twofold. First it must be decided whether the information should be made secret at all. This is done by comparing the benefits of classification to those of disclosure. Although regulations do not require the benefits of classification to be greater than the cost of classification in order to classify information, it is still a sound basis for the classification process. The costs of classification must be balanced against the costs of disclosure, which are things like detrimental effects on foreign relations or advantages gained by foreign powers to design and produce new armaments.

Against those arguments are weighed the benefits of disclosure, or in other words costs of classification. These range from the actual monetary cost of the added security to the general benefit for progress in science and technology.

At this point the information is either disclosed to public information, or it is to be classified. The security class it is assigned to is one either confidential, secret or top secret, depending on the risks it involves to national security. The following guidelines have been set in the US for classifying information:

(1) "Top Secret" shall be applied to information, the unauthorized disclosure of which reasonably could be expected to cause exceptionally grave damage to the national security.

(2) "Secret" shall be applied to information, the unauthorized disclosure of which reasonably could be expected to cause serious damage to the national security.

(3) "Confidential" shall be applied to information, the unauthorized disclosure of which reasonably could be expected to cause damage to the national security.

It is assumed that the risks involving disclosure of materials vary by an order of magnitude between the different classes. That way the damage caused by divulgence of top secret information is estimated to cause about ten times more damage than secret information, and so on. For example plans for overall conduct of war are to be classified top secret with the current system. For another example, information about strength of troops engaged in hostilities should be considered secret. Lastly, an example of confidential information would be battle reports that have value to the enemy.

In summary the analysis of costs vs. benefits determines if the information classified at all, and the nature of the information determines which security class it is assigned to. Although the real process is done in two distinct parts (decision of classification vs. disclosure and then classification to three classes), from analytical point of view it is only one process with four possible classes: disclosed, confidential, secret and top secret.
3.2.4 Downgrading and Declassifying Information

Important part of the classification system is declassification. Just like information cannot be kept perfectly safe, it cannot be kept secret for unlimited time. Even if it could, it would not be useful in all cases. For example, the time and place for invasion of Normandy was naturally Top Secret before the event, but became public when it was carried out. As another example all nuclear bomb related information was kept secret during the second world war, even information whether the bomb could be built. That information was then very loudly made public information in Hiroshima when the first atom bomb was detonated in Japan. Another factor is what the system was designed to prevent: deliberate or inadvertent disclosure of information. These types of disclosures do not necessarily mean that the information is declassified, however, much of the incentives for keeping it a secret may have vanished.

Classification of documents is thus a dynamic system. The documents have a life cycle – starting from classification to a certain class, and then they can be downgraded from one class to a lower one, upgraded if situation warrants and then finally declassified. Usually the older the classification is, the less reason there is for keeping it secret. To make the whole process optimal regarding costs involved, one must weight the costs of declassifying the document in question vs. the advantages of having it made public information. These same considerations must be taken when downgrading security status, but then comparing the reduced costs to increased risk of disclosure. Because there are significant costs involved even in reviewing the material due to number of expert personnel involved, it is not viable to keep the system “up-to-date” constantly. Rather, the documents are reviewed on certain intervals, or in rare cases they are classified for a predetermined number of years.

3.3 EBRCM – Experience Based Reliability Centered Maintenance

3.3.1 Overview

The main goal of utilities operating NPPs (Nuclear Power Plants) is economic and continued production of electricity in the long term. Safety is of utmost importance in gaining and maintaining public trust which is essential for continued operation of the plant. Maintenance of nuclear power plants has been recognized as an important factor in both economic and safe operation of nuclear power plants. Classification can be used as an aid in maintenance planning and improvement. Experience Based Reliability Centred Maintenance is one method that in part uses classification of NPP equipment to achieve maintenance goals. The advantages of classification are two-fold: equipment places numbering in the tens of thousands become manageable and it is easier to analyze and optimize maintenance actions.

Experience Based Reliability Centred Maintenance (EBRCM) uses classification to rate the significance of the equipment in question. Basically the system allows more expensive maintenance tasks for more important equipment, so that the Burden-to-Importance Ratio is steered towards 1 for all equipment. Another function of EBRCM is recognize efficient maintenance actions for different equipments by analyzing fault and maintenance history of the NPP.
3.3.2 Maintenance Strategy Classification

The importance of the equipment in question is essential in deciding what maintenance strategy is adopted for the equipment or equipment group. Possible maintenance strategies could be for example “always functioning”, “limited unavailability allowed”, “economically justified preventive maintenance” and “planned repair” in a four class system. Strategy classifications with three classes are also used. Always functioning would be the first class where the most important equipment is placed, and planned repair is the fourth and last class where equipment which do not affect plant usage are placed. Specific maintenance tasks can then be directed to the equipment so that the maintenance strategy associated with the class is fulfilled. Personnel expertise plays a large part in identification of the right, or most effective, maintenance actions. The following picture presents a decision tree for assigning maintenance classes.

![Figure 3.1: Logic Decision Tree for maintenance classes [10].](image)

The maintenance actions could consist of, for example, a mix of preventive and predictive tasks in the first class and just repairs when broken in the fourth class. Generally equipment in the fourth class do not affect plant availability. Decision trees like figure 3.1 need to be constructed according to maintenance goals at the NPP. Another, different, logic tree might be used together with maintenance experts to decide the specific maintenance tasks used to achieve the goals of the maintenance strategy.

Maintenance strategy classification gives guidelines on how actions and resources should prioritized and allocated to different systems and components in order to achieve the maintenance objectives. The logic decision tree in figure 3.1 is based on qualitative assessments but the underlying principle is in accord with the burden-to-importance ratio analysis. More important systems are given proportionally more resources for their maintenance.
4 Optimal Classification Example

4.1 Basics

We now return to the inmate classification problem presented in section 3.1. This section presents a simplified solution to that classification problem. The model used for optimization is modified from the general resource allocation model of section 2.4. The objectives of resource minimization and risk minimization are separated, so that the optimal classification problem is a multi-objectional optimization problem.

The goals decided for the prison system must be accomplished by this classification system as well as possible. The goals decided upon by US judicial system were:

- Safety of inmates and staff
- Public safety
- Protection against liability and inmates rights

These goals must be expressed mathematically to ensure that the optimization accomplishes what it is supposed. To simplify the model a risk measure is chosen for evaluating the “cost” each inmate causes in regards to risk. Another cost in the system is the resources that are used for each inmate. Efficient classification then minimizes these costs. In this simplified model the goals stated above have been all summarized in a general risk function based on the inmates attributes, and the resources the system consumes, creating a problem with just two objectives.

4.2 Constructing the Model

To formulate the problem in mathematical terms, we first need to asses the possible risks caused by the inmate. From the prison records we could identify common or particularly harmful crimes, and use statistical analysis to calculate how the attributes from the criminals data affect chances to being committing different crimes. A trickier part is to assign a negative “worth” for the crime. Let us assume that officials are capable of putting a cost for each crime, so that we will have a probability and cost for each crime based on the criminals attributes and the security class.

<table>
<thead>
<tr>
<th>No.</th>
<th>Scenario</th>
<th>Probability</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Riot</td>
<td>$P_1(x,c)$</td>
<td>$D_1$</td>
</tr>
<tr>
<td>2</td>
<td>Violence</td>
<td>$P_2(x,c)$</td>
<td>$D_2$</td>
</tr>
<tr>
<td>3</td>
<td>Drugs</td>
<td>$P_3(x,c)$</td>
<td>$D_3$</td>
</tr>
<tr>
<td>4</td>
<td>Minor infraction</td>
<td>$P_4(x,c)$</td>
<td>$D_4$</td>
</tr>
<tr>
<td>5</td>
<td>No crime</td>
<td>$P_5(x,c) = 1-\sum P_i(x,c)$</td>
<td>$D_5 = 0$</td>
</tr>
</tbody>
</table>

Table 1 Examples of actions constituting risk in prisons

The risk posed by each inmate can now be calculated as an expected value:
\[ f(x,c) = \sum_{k=1}^{4} P_k(x,c)D_k \]  

(4.1)

Where \( P \) is the probability of each infraction and \( D \) is the respective cost. Only four possible crimes are considered, to simplify the example. \( P \) depends on the prisoners attributes \( x \) and \( c \), the security level of the prison he is assigned into. \( f(x,c) \) is the prisoners risk function. The function is discrete, specified only in three points – the three security classes. Different prisoners risk functions might look like the following:

![Inmate Risk Functions](image)

**Figure 4.1 Examples of possible risk functions for inmates**

Figure 4.1 illustrates the fact that the risk posed by the inmates can depend on the security level of the prison in different, non-linear, ways. The example inmates are dangerous, aggressive and benevolent.

In designing the prisoner classification we are aiming for a safe and efficient system, and to meet these ends two criteria were chosen to be minimized: overall risk of the system, and the resources used. Optimal classification problem can then be written as:

minimize \[ \bar{P}(\bar{x}, \bar{c}) = \begin{bmatrix} F \\ C \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} f(x_i, c_i) \\ \sum_{i=1}^{N} s(c_i) \end{bmatrix} \]  

(4.2)

with respect to \( \bar{c} = (c_1, ..., c_N) \), \( c_i \in \{1,2,3\} \)

subject to \[ \sum_{i=1}^{N} s(c_i) \leq R \]  

(4.3)

\[ \sum_{i=1}^{N} f(x_i, c_i) \leq F \]  

(4.4)

and \[ \sum_{i=1}^{N} 1_{\{c_i=3\}} \leq T_j \]  

(4.5)
where $N$ is the number of inmates to be classified, $V$ is the target function which is two-dimensional with both the total risk and the total resources as objectives, $f(x, c)$ is the risk function as defined in (4.1), $x_i$ is the attributes of the prisoner, $c_i \in \{1,2,3\}$ is one of the three possible security classes inmate $i$ is assigned into, $s(c)$ is the function portraying the loss of resources single inmate causes in security level $c$, $R$ is the amount of total resources available in the prison system and $T_j$ is the number of inmates security level $j$ of the prison can accommodate. Function $s(c_j)$ is defined as the following in this example: $s(1)=s_1$, $s(2)=s_2$ and $s(3)=s_3$. It is assumed that the resources consumed are greater in a higher class so $s_1 < s_2 < s_3$. Correspondingly, the risk contribution is decreasing: $f(x, 1) > f(x, 2) > f(x, 3)$.

The limits $R$, $F$ and $T_j$ are imposed for the resources used, overall risk and the inmate capability in security class $j$, respectively. Figure 4.2 illustrates these limits.

### 4.3 Optimization

#### 4.3.1 Multi-objective Optimization

The optimization problem is $N$-variable discrete multi-objective optimization problem where variables $\{c_i\}_{i=1}^{N}$ may assume one of the three values 1, 2 or 3, each corresponding to one of the security levels. The two objectives to be minimized are the overall risk in the system and the resources spent to counter that risk. The costs associated with classes 1, 2 and 3 are $s_1$, $s_2$ and $s_3$, irrelevant of the attributes of the inmate. In a multi-objective problem such as this the objective is to find a pareto-optimal surface or, in this two-dimensional case, a curve.

The following figure 4.2 illustrates the optimization problem. The graph shows the efficient border for risk vs. resources spent; i.e. it shows the pareto-optimal curve. Above that curve both resources and the overall risk can be reduced, but on the curve either one must be increased if the other one is decreased. It follows that we are mainly interested how to calculate the points on the pareto-optimal curve. Arrows indicate change in pareto-optimal curve when an item is added to the system, with each of the three arrows corresponding to one of three classes. The arrows are a graphical presentation of the risk functions and resource costs of each inmate, and thus possibly different for each inmate. Because of the different risk functions (4.1) for each inmate the empirical $F(C)$ function would not be as smooth as shown in figure 4.2, nor is it even continuous. Limits for risk and resources are included.

![Figure 4.2 Risk vs. resources efficient border](image-url)
4.3.2 Calculating the Pareto-optimal Curve

The main purpose of the model is to make the classification system as efficient as possible. We want to find those classification vectors that have such \( F \) and \( C \) that they both can not be reduced by reclassifying any of the items. These \( F \) and \( C \) pairs constitute the pareto-optimal curve. When that curve is constructed it is just a matter of decision analysis which of those pairs is chosen for the actual classification scheme.

In this example there are 3 possible classes, and thus a set \( 3^N \) possible solutions for overall classification. A brute force approach would be eliminate non-pareto-optimal solutions from this set, but as \( N \) increases, this may prove too time-consuming a task. A bit more elegant solution is to add items to the system one at a time in iterative fashion. Let us use the following notations:

\[
A_k \quad \text{Is the set of all possible classifications for } k \text{ items, such that } A_k = D_1 \times \ldots \times D_k, \text{ where } D_i \text{ is the class of component number } i, \text{ so that } D = \{1,2,3\}.
\]

\[
E_k \quad \text{Is the set of all efficient classifications for } k \text{ item, or in other words the pareto-optimal set for total risk } F \text{ versus total resources used } C.
\]

\[
H_k \quad \text{Is the set of all non-optimal classification solutions, defined as } H^k = \left(A^k \setminus E^k \right).
\]

Now the new set of pareto-optimal solutions with \( k+1 \) items \( E_{k+1} \) can be calculated by adding one new component to the pareto-optimal solution with \( k \) items. Thus getting a new set of solutions:

\[
E'_{k+1} = E_k \times D
\]

\( E'_{k+1} \) is the set of candidates for a pareto-optimal solutions. From this set the non-pareto-optimal are eliminated to find \( E_{k+1} \). In appendix B it is proven that an addition of item to a non-pareto-optimal solution with \( k \) items cannot yield a pareto-optimal solution with \( k+1 \) items, thus it applies for the system with \( k+1 \) items:

\[
E_{k+1} \cap H'_{k+1} = \phi,
\]

where \( H'_{k+1} = H_k \times D \).

So all pareto-optimal solutions will be found with this iterative method.

4.3.3 Classifying New Components

Calculating the pareto-optimal curve can be time-consuming when there is a large number of classifiable items in the system. A new curve would have to be calculated each time new
component was added. It is possible, however, to choose the most efficient class for a new item brought into the system provided that a few assumptions are made.

The assumptions are that the total risk $F$ and total amount of resources spent $C$ are large compared to the risk caused by single component $f(x,c)$ and the resources consumed by a single component $s(c)$. $F(C)$ is a decreasing function and that $s(1) < s(2) < s(3)$ and $f(x,1) > f(x,2) > f(x,3)$.

The basis for this quicker method is the examination of changes the new component causes in the pareto-optimal curve. Pareto-optimal curve is the curve defined by minimum possible risk at each amount of resources spent. The three possible classifications for the new item mean that at each point there are three new possible pareto-optimal curves. At each point we simply assign the new item to the class that has the candidate curve closest to the original pareto-optimal curve.

Terms used:

- $F(C)$ = Polynomial approximation of the set of pareto-optimal points
- $f(x,c)$ = Risk increase of single item added to the system
- $s(c)$ = Resources consumed by a single item in class c

When item $x$ classified to class 1 is added to the system the new curve is:

$$F_1(C) = F(C - s(1)) + f(x,1)$$  \hspace{1cm} (4.7)

In the same way the equation for the curves for the other two classes are:

$$F_2(C) = F(C - s(2)) + f(x,2)$$  \hspace{1cm} (4.8)

$$F_3(C) = F(C - s(3)) + f(x,3)$$  \hspace{1cm} (4.9)

Since $F$ is decreasing, $F_1(C)$, $F_2(C)$ and $F_3(C)$ are decreasing.

To choose the most efficient class for the new component we will have to evaluate which of the three curves is closest to the original pareto-optimal curve (which does not include the new item at all.) The new curve, $F'(C)$ is

$$F'(C) = \min\{F_1(C), F_2(C), F_3(C)\}.$$  \hspace{1cm} (4.10)

This is now the new pareto-optimal curve after the addition of one item, because it has the minimum total risk at each $C$.

The new item is then assigned to that class which has the lowest value according to (4.10). The following figure illustrates two possibilities when adding a new component to the system. The two cases a) and b) can be thought of as close-ups of figure 4.2 at different values of $C$. 

23
Because it was assumed that $F$ and $C$ were large compared to $f(x,c)$ and $s(c)$, and that $F(c)$ is polynomial approximation of the discrete pareto-optimal curve, it follows that $F’(C)$ can be assumed linear when evaluating the impact of a single new component. In general $F(C)$ would look something like in figure 4.2. In situation a. in figure 4.3 the second class would be chosen, but in b. class 1 would chosen.

### 4.3.4 The Role of the Middle Alternative

Some of the classification tasks examined throughout section 2 used classification schemes with more than two classes. It is thus interesting to examine what mechanics factor into the number of classes that are desirable. A case with two classes is not as interesting because depending on the decision makers’ preferred $F/C$ ratio one of the two classes will always be chosen. With three classes it is possible for the middle alternative to be obsolete, or non-optimal choice for all $F/C$ ratios. For that purpose a classification scheme with three classes is examined here, in order to find out which assumptions lead to middle class being obsolete.

In this example with three classes the second (middle) class is not a viable option for every inmate. We will assume that $F$ and $C$ are large compared to $f(x,c)$ and $s(c)$. Then it can be proven that only if the risk function of the inmate (or in other words the inmates response to added security) is convex, the second class will prove to be viable at some level of use of resources $C$. This is proven in Appendix C. Figure 4.4 illustrates the effects of convex and concave risk functions on the pareto-optimal curve.
The shape of the risk functions of single inmates is also essential in deciding how many classes are required for efficient classification system. If most of the risk functions were concave, there would be no need for a second class at all. If enough individuals have a convex function a second class is appropriate.

If a two class system were used instead of three the risk will be needlessly increased in case the inmates risk function is convex. Figure 4.5 illustrates this. The three different curves above the lowest, pareto-optimal curve, represent the additions to risk for each class the new inmate could be assigned into. As can be seen in figure 4.5 the second class allows for lower total risk with some values of resource C.
4.3.5 Combining Optimization Objectives

There are different approaches to solving a multi-objective optimization problem. Weighting method is one that can be used. Each of the objectives is multiplied with a weight value $0 \leq W_i \leq 1$, and the modified functions are then summed up to form a single-objective problem. The weighted target function would in this case be:

$$ V(\bar{x}, \bar{c}) = W_1 \sum_{i=1}^{N} s(c_i) + W_2 \sum_{i=1}^{N} f(x_i, c_i) $$  \hspace{1cm} (3.6)

With the additional constraints:

$$ W_1 + W_2 = 1 $$
$$ W_1, W_2 \geq 0 $$  \hspace{1cm} (3.7)

Target function is the same now as presented in section 2. Decision analysis should be used to determine $W_1$ and $W_2$ with the people who are participating in designing the system. The weights used will determine where the solution is in the pareto-optimal curve depicted in figure 4.2.

Any number of other decision theory methods could be used to to choose one of the efficient solutions calculated in 4.3.1.
5 Discussion and Conclusions

Classification tasks are an efficient way to improve many different real world systems, as presented in section 2. Intuitively applied classifications can help people understand and manage complex systems, in situations where micro-management or observation of single items is not possible. In these kinds of cases mathematical analysis of the problem is not always needed, or it need not be very thorough. An example of this might be a vacation brochure – allocating hotels to different classes is not actually a mathematical science, even though the stars do give a good overview of the hotels services.

On another level classification can be used as a tool in systems analysis. In this kind of cases mathematics are more important, as the goal usually is optimal functionality. Classification task applied correctly will allow us to use simpler mathematics than a case with no classification implemented, when optimizing the system. This is analogous to the more intuitive cases in a sense, because classification should make the system simpler in both cases – to understand it in the other, and mathematically simpler in the other.

A third kind of case might be such that the classification system is forced, or the only way to do a certain thing. This is common in the sense that if the classes correspond to some real world objects, there are not infinite numbers of those. Like the prison security levels in section 4 – one can not have custom built prisons for each prisoner. In this sense the classification is forced to a few security levels, but decision makers are allowed to tinker with those levels.

It must be noted that a good mathematical model of the system and a clear picture of the goals for the classification task have to be considered when constructing a classification system. This study did not concentrate on constructing the overall classification process too much (except for section 4.3.4), but these two rules must be observed so that the classification itself does not become the end instead of means.
Appendix A

Proof of BIR-measure

Let us prove that the BIR-measure approaches unity at optimal point. Utility function is calculated with weighted combination of the total risk $F$ and total resources used $C$, in the following way:

$$V = \alpha F + (1-\alpha)C \quad (A.1)$$

At optimum allocations the following condition must apply:

$$\frac{\partial V}{\partial c_i} = \alpha \frac{\partial F}{\partial c_i} + (1-\alpha) \frac{\partial C}{\partial c_i} = 0 \quad (A.2)$$

Or otherwise:

$$\alpha \frac{\partial F}{\partial c_i} = -(1-\alpha) \frac{\partial C}{\partial c_i} \quad (A.3)$$

Summing equation (A.3) over $n$ variables gives us:

$$\alpha \sum_{i=1}^{n} \frac{\partial F}{\partial c_i} = -(1-\alpha) \sum_{i=1}^{n} \frac{\partial C}{\partial c_i} \quad (A.4)$$

Dividing equation (A.3) with equation (A.4) yields

$$\frac{\alpha \sum_{i=1}^{n} \frac{\partial F}{\partial c_i}}{\alpha \sum_{i=1}^{n} \frac{\partial F}{\partial c_i}} = \frac{-(1-\alpha) \sum_{i=1}^{n} \frac{\partial C}{\partial c_i}}{-(1-\alpha) \sum_{i=1}^{n} \frac{\partial C}{\partial c_i}} \quad (A.5)$$

The weighting factors $\alpha$ and $(1-\alpha)$ are divided out. Next we will divide the right-hand side with the left-hand side.

$$\frac{\partial C}{\partial c_i} / \sum_{i=1}^{n} \frac{\partial C}{\partial c_i} = 1$$

$$\frac{\partial F}{\partial c_i} / \sum_{i=1}^{n} \frac{\partial F}{\partial c_i} = 1$$

This shows that at optimal allocation the fractional increase in resources $C$ divided by the fractional increase in risk equals to one.
Appendix B

Constructing the Pareto-optimal Set

Let \( A_k \) be the set of all possible classifications for \( k \) item, such that \( A_k = D_1 \times \ldots \times D_k \), where \( D_i \) is the class of component number \( i \), so that \( D = \{1,2,3\} \).

\( E_k \) is the set of all efficient classifications for \( k \) item, or in other words the pareto-optimal set for total risk \( F \) versus total resources used \( C \).

\( H_k \) is the set of all non-optimal classification solutions, defined as \( H^k = \left( A^k \setminus E^k \right) \).

The point of this proof is to show that the series of pareto-optimal solutions \( \{E_j\}_{j=1}^{\infty} \) can be constructed by adding items to the system one at a time:

\[
E'_{k+1} = E_k \times D
\]  (B.1)

Where \( E'_{k+1} \) is a set of possible pareto-optimal points, from which the non-optimal solutions are eliminated by comparing the total risk function and resource values in the way explained in section 4.2, using the following formulation for the points in set \( H_k \) :  

\[
\text{if } e_j \in H_k, \text{ then } \exists e_j \in A_k \text{ so that both } F(e_j) < F(e_i) \text{ and } C(e_j) < C(e_i) \]  (B.2)

We also want to maintain that the construction of optimal solutions is unique, meaning that an optimal solution can not be achieved by adding an item to a non-optimal solution. In formal way this means that:

\[
H'_{k+1} \cap E_{k+1} = \emptyset \]  (B.3)

where \( H'_{k+1} \) is constructed like:

\[
H'_{k+1} = H_k \times D = E^C_k \times D
\]  (B.4)

Let us mark a non-pareto-optimal solution with \( e_i \in H_k \). Then there exists a solution \( e_j \in E_k \), which is pareto-optimal, such that:

\[
F(e_j) < F(e_i), \text{ and } C(e_j) < C(e_i)
\]  (B.5)
If (B.5) was not true, \( e_i \) would be pareto-optimal by definition (B.2). Because of the additive risk model used, it follows that for all \( e'_{ij} \in H'_{k+1} \) there exists a dominating solution \( e'_{ij} \in E'_{k+1} \) derived from the pareto-optimal solution will satisfy the following for each \( D = \{1,2,3\} \):

\[
F(e'_{ij}) = F(e_i) + \Delta F(D) > F(e_j) + \Delta F(D) = F(e'_{ij}), \text{ and}
\]
\[
C(e'_{ij}) = C(e_i) + \Delta C(D) > C(e_j) + \Delta C(D) = C(e'_{ij}),
\]
(B.6)

Thus \( e'_{ij} \in H_{k+1} \) can not be a pareto-optimal solution according to (B.2).

In simpler terms, this means that for whatever class the new item is assigned, the result will be less if it is assigned to the same class but on the pareto-optimal solution, instead of the non-pareto-optimal solution.

This means that

\[
H'_{k+1} \cap E_{k+1} = \emptyset.
\]
(B.7)
Appendix C

Proof for Viability of Middle Class

In section 4.3.4 it was noted that only a convex risk individual function $f(x,c)$ results in a possible allocation to middle class, or in this case class 2. While the risk function $f(x,c)$ is discrete because it is only defined at three points, we will define risk function as continuous here with a notation $f(s(c))$ (it is worth noting that the cost $s(c)$ is still defined only at three points). For a convex function we know:

$$f((1 - \lambda)s(1)) + \lambda f(s(3)) < (1 - \lambda)f(s(1)) + \lambda f(s(3)) \quad \text{(C.1)}$$

With that assumption we will prove that at least one of the following inequalities holds true:

$$F_1(C) < F_2(C) \quad \text{(C.2 a and b)}$$
$$F_3(C) < F_2(C)$$

If we can prove that, we have shown that $F_2(C)$ is never pareto-optimal, and thus the new item is never classified to class 2.

To ease the calculation, let’s introduce a measure that portrays the vertical difference between the pareto-optimal curve and the curve constructed by adding one item to class $i$: (C.2):

$$d_i = f(s(i)) - s(i)F'(C) \quad \text{(C.3)}$$

The next picture illustrates this measure.

![Figure C.1 Figure of distance measure $d_i$](image)
It is clear that (C.2) is equivalent with:

\[
\begin{align*}
  d_1 &< d_2 \\
  d_3 &< d_4 \\
\end{align*}
\]  
(C.4a and b)

It is thus sufficient to prove that (C.4) holds when \( f(s(i)) \) is concave. When \( f(s(i)) \) is concave:

\[
\begin{align*}
  f(s(2)) &> (1 - \lambda) f(s(1)) + \lambda f(s(3)) \\
  \text{where } \lambda &= \frac{s(2) - s(1)}{s(3) - s(1)}, \\
\end{align*}
\]  
(C.5)

so that \( s(2) = (1 - \lambda)s(1) + \lambda s(3) \). Then we will combine this with (C.3) and use \( f(s(i)) \)'s concavity in (C.5) and (C.6) to get:

\[
\begin{align*}
  d_2 &= f(s(2)) - s(2)F'(C) \\
  &= f(s(2)) - ((1 - \lambda)s(1) + \lambda s(3))F'(C) \\
  &= (1 - \lambda) f(s(1)) + \lambda f(s(3)) - ((1 - \lambda)s(1) + \lambda s(3))F'(C) \\
  &= (1 - \lambda)(f(s(1)) - s(1)F'(C)) + \lambda(f(s(3)) - s(3)F'(C)) \\
  &= d_1 + d_3 \\
  &> \min(d_1, d_3). \\
\end{align*}
\]  
(C.6)

This is true only if \( d_2 > d_1 \) or \( d_2 > d_3 \). Thus the proof is complete.
References


