

Machine Learning
DIVY PATEL 171080056
Under the guidance of Pranav Sir

Thesis in Information Technology

April 10, 2019

University of Jyväskylä

Faculty of Information Technology

Author: *Did you remember to use the `\author` command?*

Contact information: *Did you remember to use the `\contactinformation` command?*

Supervisor: Supervised work

Title: Machine Learning DIVY PATEL 171080056 Under the guidance of Pranav Sir

Työn nimi: L^AT_EX-tutkielmapohjan gradu3 käyttö

Project: Thesis

Study line: All study lines

Page count: 16+2

Abstract: This document is a sample gradu3 thesis document class document. It also functions as a user manual and supplies guidelines for structuring a thesis document.

The abstract is typically short and discusses the background, the aims, the research methods, the obtained results, the interpretation of the results and the conclusions of the thesis. It should be so short that it, the Finnish translation, and all other meta information fit on the same page.

The Finnish tiivistelmä of a thesis should usually say exactly the same things as the abstract.

Keywords: L^AT_EX, gradu3, Master's Theses, Bachelor's Theses, user's guide

Suomenkielinen tiivistelmä: Tämä kirjoitelma on esimerkki siitä, kuinka gradu3-tutkielmapohjaa käytetään. Se sisältää myös käyttöohjeet ja tutkielman rakennetta koskevia ohjeita.

Tutkielman tiivistelmä on tyypillisesti lyhyt esitys, jossa kerrotaan tutkielman taustoista, tavoitteesta, tutkimusmenetelmistä, saavutetuista tuloksista, tulosten tulkinnasta ja johtopäätöksistä. Tiivistelmän tulee olla niin lyhyt, että se, englanninkielinen abstrakti ja muut metatiedot mahtuvat kaikki samalle sivulle.

Sen tulee kertoa täsmälleen samat asiat kuin englanninkielinen abstrakti.

Avainsanat: L^AT_EX, gradu3, pro gradu -tutkielmat, kandidaatintutkielmat, käyttöohje

Preface

This is where you can write a preface for your thesis. Most theses don't have prefaces, but if you write one, keep it short (at least one page).

The preface should discuss more the thesis process than the content of the thesis. For example, if there is something out of the ordinary in your choice of a thesis topic or if something out of the ordinary happened during its preparation, the preface is where you could write about it. It is also customary in a preface to thank by name those persons who helped you with your thesis – at least your supervisor, your spouse and your children, if any. (Your family likely will have helped you by encouraging and supporting you.)

The preface is typically in the first person ("I"). It is also common to sign it. April 10, 2019

The Author

Glossary

$\text{T}_{\text{E}}\text{X}$	A batch-oriented typesetting system written by Donald Knuth in 1977–1989 (see Knuth 1986).
\LaTeX	A system, built on top of $\text{T}_{\text{E}}\text{X}$ (Knuth 1986), for typesetting structured documents (see Lamport 1994). Its current version is $\text{\LaTeX} 2_{\epsilon}$.

List of Tables

Table 1. Command changes from gradu2 to gradu3	11
--	----

Contents

1	INTRODUCTION	1
2	MACHINE LEARNING TASKS	2
3	HISTORY AND RELATIONSHIPS TO OTHER FIELDS.....	4
3.1	Relation to Data Mining	5
3.2	Relation to optimisation.	5
3.3	Relation to statistics	6
4	THEORY	7
5	APPROACHES	8
5.1	Supervised and Semi-Supervised Learning.	8
5.2	Unsupervised Learning.....	8
5.3	Reinforcement Learning	9
	BIBLIOGRAPHY	10
	APPENDICES	11
A	Moving from gradu2 to gradu3	11
B	Rarely needed features	12

1 Introduction

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to effectively perform a specific task without using explicit instructions, relying on patterns and inference instead.

It is seen as a subset of artificial intelligence. Machine learning algorithms build a mathematical model of sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task.[1][2]:2 Machine learning algorithms are used in the applications of email filtering, detection of network intruders, and computer vision, where it is infeasible to develop an algorithm of specific instructions for performing the task. Machine learning is closely related to computational statistics, which focuses on making predictions using computers. The study of mathematical optimization delivers methods, theory and application domains to the field of machine learning. Data mining is a field of study within machine learning, and focuses on exploratory data analysis through unsupervised learning.[3][4] In its application across business problems, machine learning is also referred to as predictive analytics.

The name machine learning was coined in 1959 by Arthur Samuel.[5] Tom M. Mitchell provided a widely quoted, more formal definition of the algorithms studied in the machine learning field: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T , as measured by P , improves with experience E ."[6] This definition of the tasks in which machine learning is concerned offers a fundamentally operational definition rather than defining the field in cognitive terms. This follows Alan Turing's proposal in his paper "Computing Machinery and Intelligence", in which the question "Can machines think?" is replaced with the question "Can machines do what we (as thinking entities) can do?".[7] In Turing's proposal the various characteristics that could be possessed by a thinking machine and the various implications in constructing one are exposed.

2 Machine Learning Tasks

Machine learning tasks are classified into several broad categories. In supervised learning, the algorithm builds a mathematical model from a set of data that contains both the inputs and the desired outputs. For example, if the task were determining whether an image contained a certain object, the training data for a supervised learning algorithm would include images with and without that object (the input), and each image would have a label (the output) designating whether it contained the object. In special cases, the input may be only partially available, or restricted to special feedback.[clarification needed] Semi-supervised learning algorithms develop mathematical models from incomplete training data, where a portion of the sample input doesn't have labels.

Classification algorithms and regression algorithms are types of supervised learning. Classification algorithms are used when the outputs are restricted to a limited set of values. For a classification algorithm that filters emails, the input would be an incoming email, and the output would be the name of the folder in which to file the email. For an algorithm that identifies spam emails, the output would be the prediction of either "spam" or "not spam", represented by the Boolean values true and false. Regression algorithms are named for their continuous outputs, meaning they may have any value within a range. Examples of a continuous value are the temperature, length, or price of an object.

In unsupervised learning, the algorithm builds a mathematical model from a set of data which contains only inputs and no desired output labels. Unsupervised learning algorithms are used to find structure in the data, like grouping or clustering of data points. Unsupervised learning can discover patterns in the data, and can group the inputs into categories, as in feature learning. Dimensionality reduction is the process of reducing the number of "features", or inputs, in a set of data.

Active learning algorithms access the desired outputs (training labels) for a limited set of inputs based on a budget, and optimize the choice of inputs for which it will acquire training labels. When used interactively, these can be presented to a human user for labeling. Reinforcement learning algorithms are given feedback in the form of positive or negative

reinforcement in a dynamic environment, and are used in autonomous vehicles or in learning to play a game against a human opponent.[2]:3 Other specialized algorithms in machine learning include topic modeling, where the computer program is given a set of natural language documents and finds other documents that cover similar topics. Machine learning algorithms can be used to find the unobservable probability density function in density estimation problems. Meta learning algorithms learn their own inductive bias based on previous experience. In developmental robotics, robot learning algorithms generate their own sequences of learning experiences, also known as a curriculum, to cumulatively acquire new skills through self-guided exploration and social interaction with humans. These robots use guidance mechanisms such as active learning, maturation, motor synergies, and imitation.[clarification needed]

3 History and relationships to other fields.

Arthur Samuel, an American pioneer in the field of computer gaming and artificial intelligence, coined the term "Machine Learning" in 1959 while at IBM[8]. As a scientific endeavour, machine learning grew out of the quest for artificial intelligence. Already in the early days of AI as an academic discipline, some researchers were interested in having machines learn from data. They attempted to approach the problem with various symbolic methods, as well as what were then termed "neural networks"; these were mostly perceptrons and other models that were later found to be reinventions of the generalized linear models of statistics.[9] Probabilistic reasoning was also employed, especially in automated medical diagnosis.[10]:488

However, an increasing emphasis on the logical, knowledge-based approach caused a rift between AI and machine learning. Probabilistic systems were plagued by theoretical and practical problems of data acquisition and representation.[10]:488 By 1980, expert systems had come to dominate AI, and statistics was out of favor.[11] Work on symbolic/knowledge-based learning did continue within AI, leading to inductive logic programming, but the more statistical line of research was now outside the field of AI proper, in pattern recognition and information retrieval.[10]:708–710; 755 Neural networks research had been abandoned by AI and computer science around the same time. This line, too, was continued outside the AI/CS field, as "connectionism", by researchers from other disciplines including Hopfield, Rumelhart and Hinton. Their main success came in the mid-1980s with the reinvention of backpropagation.[10]:25

Machine learning, reorganized as a separate field, started to flourish in the 1990s. The field changed its goal from achieving artificial intelligence to tackling solvable problems of a practical nature. It shifted focus away from the symbolic approaches it had inherited from AI, and toward methods and models borrowed from statistics and probability theory.[11] It also benefited from the increasing availability of digitized information, and the ability to distribute it via the Internet.

3.1 Relation to Data Mining

Machine learning and data mining often employ the same methods and overlap significantly, but while machine learning focuses on prediction, based on known properties learned from the training data, data mining focuses on the discovery of (previously) unknown properties in the data (this is the analysis step of knowledge discovery in databases). Data mining uses many machine learning methods, but with different goals; on the other hand, machine learning also employs data mining methods as "unsupervised learning" or as a preprocessing step to improve learner accuracy. Much of the confusion between these two research communities (which do often have separate conferences and separate journals, ECML PKDD being a major exception) comes from the basic assumptions they work with: in machine learning, performance is usually evaluated with respect to the ability to reproduce known knowledge, while in knowledge discovery and data mining (KDD) the key task is the discovery of previously unknown knowledge. Evaluated with respect to known knowledge, an uninformed (unsupervised) method will easily be outperformed by other supervised methods, while in a typical KDD task, supervised methods cannot be used due to the unavailability of training data.

3.2 Relation to optimisation.

Machine learning also has intimate ties to optimization: many learning problems are formulated as minimization of some loss function on a training set of examples. Loss functions express the discrepancy between the predictions of the model being trained and the actual problem instances (for example, in classification, one wants to assign a label to instances, and models are trained to correctly predict the pre-assigned labels of a set of examples). The difference between the two fields arises from the goal of generalization: while optimization algorithms can minimize the loss on a training set, machine learning is concerned with minimizing the loss on unseen samples.[12]

3.3 Relation to statistics

Machine learning and statistics are closely related fields. According to Michael I. Jordan, the ideas of machine learning, from methodological principles to theoretical tools, have had a long pre-history in statistics.[13] He also suggested the term data science as a placeholder to call the overall field.[13]

Leo Breiman distinguished two statistical modelling paradigms: data model and algorithmic model,[14] wherein "algorithmic model" means more or less the machine learning algorithms like Random forest.

Some statisticians have adopted methods from machine learning, leading to a combined field that they call statistical learning.[15] On the command line, biber is simple to use. Once \LaTeX (or pdf\LaTeX) has been run once, invoke biber with the document name (without the .tex part) as its argument. After that, run \LaTeX (or pdf\LaTeX) at least once, until the latest run does not request another run. For example:

4 Theory

A core objective of a learner is to generalize from its experience.[2][16] Generalization in this context is the ability of a learning machine to perform accurately on new, unseen examples/tasks after having experienced a learning data set. The training examples come from some generally unknown probability distribution (considered representative of the space of occurrences) and the learner has to build a general model about this space that enables it to produce sufficiently accurate predictions in new cases.

The computational analysis of machine learning algorithms and their performance is a branch of theoretical computer science known as computational learning theory. Because training sets are finite and the future is uncertain, learning theory usually does not yield guarantees of the performance of algorithms. Instead, probabilistic bounds on the performance are quite common. The bias–variance decomposition is one way to quantify generalization error.

For the best performance in the context of generalization, the complexity of the hypothesis should match the complexity of the function underlying the data. If the hypothesis is less complex than the function, then the model has underfit the data. If the complexity of the model is increased in response, then the training error decreases. But if the hypothesis is too complex, then the model is subject to overfitting and generalization will be poorer.[17]

In addition to performance bounds, computational learning theorists study the time complexity and feasibility of learning. In computational learning theory, a computation is considered feasible if it can be done in polynomial time. There are two kinds of time complexity results. Positive results show that a certain class of functions can be learned in polynomial time. Negative results show that certain classes cannot be learned in polynomial time.

5 Approaches

5.1 Supervised and Semi-Supervised Learning.

Supervised learning algorithms build a mathematical model of a set of data that contains both the inputs and the desired outputs.[18] The data is known as training data, and consists of a set of training examples. Each training example has one or more inputs and a desired output, also known as a supervisory signal. In the case of semi-supervised learning algorithms, some of the training examples are missing the desired output. In the mathematical model, each training example is represented by an array or vector, and the training data by a matrix. Through iterative optimization of an objective function, supervised learning algorithms learn a function that can be used to predict the output associated with new inputs.[19] An optimal function will allow the algorithm to correctly determine the output for inputs that were not a part of the training data. An algorithm that improves the accuracy of its outputs or predictions over time is said to have learned to perform that task.[6]

Supervised learning algorithms include classification and regression.[20] Classification algorithms are used when the outputs are restricted to a limited set of values, and regression algorithms are used when the outputs may have any numerical value within a range. Similarity learning is an area of supervised machine learning closely related to regression and classification, but the goal is to learn from examples using a similarity function that measures how similar or related two objects are. It has applications in ranking, recommendation systems, visual identity tracking, face verification, and speaker verification.

5.2 Unsupervised Learning.

Unsupervised learning algorithms take a set of data that contains only inputs, and find structure in the data, like grouping or clustering of data points. The algorithms therefore learn from test data that has not been labeled, classified or categorized. Instead of responding to feedback, unsupervised learning algorithms identify commonalities in the data and react based on the presence or absence of such commonalities in each new piece of data. A cen-

tral application of unsupervised learning is in the field of density estimation in statistics,[21] though unsupervised learning encompasses other domains involving summarizing and explaining data features.

Cluster analysis is the assignment of a set of observations into subsets (called clusters) so that observations within the same cluster are similar according to one or more predesignated criteria, while observations drawn from different clusters are dissimilar. Different clustering techniques make different assumptions on the structure of the data, often defined by some similarity metric and evaluated, for example, by internal compactness, or the similarity between members of the same cluster, and separation, the difference between clusters. Other methods are based on estimated density and graph connectivity.

5.3 Reinforcement Learning

Reinforcement learning is an area of machine learning concerned with how software agents ought to take actions in an environment so as to maximize some notion of cumulative reward. Due to its generality, the field is studied in many other disciplines, such as game theory, control theory, operations research, information theory, simulation-based optimization, multi-agent systems, swarm intelligence, statistics and genetic algorithms.[22][23] In machine learning, the environment is typically represented as a Markov Decision Process (MDP). Many reinforcement learning algorithms use dynamic programming techniques.[22][23][24] Reinforcement learning algorithms do not assume knowledge of an exact mathematical model of the MDP, and are used when exact models are infeasible.[22][23] Reinforcement learning algorithms are used in autonomous vehicles or in learning to play a game against a human opponent.

Bibliography

Knuth, Donald E. 1986. *The T_EXbook*. Reading, MA: Addison Wesley.

Lamport, Leslie. 1994. *L^AT_EX: A Document Preparation System*. 2nd edition. Reading, MA: Addison–Wesley.

Appendices

A Moving from gradu2 to gradu3

Moving an incomplete thesis from gradu2 to gradu3 is not particularly difficult. The first thing to do is to change gradu2 into gradu3 in the `\documentclass` command. Most of the options given to it must be removed, as they are not supported. A “kandi” option is changed into “bachelor”; any “english” option is retained, and so is “utf8”, “latin1”, or “latin9”.

Table 1 lists the command name changes that are needed. A dash indicates that there is no corresponding command. Note especially the new commands.

gradu2	gradu3
—	<code>\maketitle</code>
—	<code>\supervisor</code>
<code>\acmccs</code>	—
<code>\aine</code>	<code>\subject</code>
<code>\copyrightowner</code>	—
<code>\fulltitle</code>	—
<code>\laitos</code>	<code>\department</code>
<code>\license</code>	—
<code>\linja</code>	<code>\studyline</code>
<code>\paikka</code>	—
<code>\setauthor</code>	<code>\author</code>
<code>\termlist</code>	thetermlist environment
<code>\tyyppi</code>	<code>\type</code>
<code>\yhteystiedot</code>	<code>\contactinformation</code>
<code>\yliopisto</code>	<code>\university</code>
<code>\ysa</code>	—

Table 1. Command changes from gradu2 to gradu3

The most effort is likely needed to converting citations and the bibliography.

B Rarely needed features

In addition to features already mentioned, gradu3 offers the following additional features:

- The standard options “draft” and “final” work.
- The option “finnish” works (but is not needed, as it is the default).
- You can change the University of the thesis by using the `\university` command.
- You can change the Department of the thesis by using the `\department` command.
- You can change the formal subject of the thesis by using the `\subject` command. In English theses, the subject should be prefixed by “in” (for example, “in Information Technology”); in Finnish theses, use a capital initial letter and the genitive form (“*Tietotekniikan*”).
- You can change the type of the thesis by using the `\type` command.
- You can set the date of the thesis by using the `\setdate` command. Give it three parameters (day of month, month, and year) in numerical form.
- The `chapterquote` environment can be used to give an epigraph to a chapter. There is one mandatory parameter (the attribution of the epigraph).
- The command `\graducslsdate` prints the release date of the current version of gradu3, and the command `\graducslsversion` prints its version number.