COS4852

Machine Learning

UNIT 0 – What this module is about and where to find the material

0.1 Introduction

Machine Learning *Machine learning is the science of getting computers to act without being explicitly programmed.*

- from the Stanford online course on Machine Learning, by Andrew Ng

Machine Learning *is a scientific discipline that is concerned with the design and development of algorithms that allow computers to change behavior based on data, such as from sensor data or databases. A major focus of machine learning research is to automatically learn to recognize complex patterns and make intelligent decisions based on data.*

- from http://en.wikipedia.org/wiki/Machine_learning

The objective of this module is to provide an introduction to the theoretical and some practical aspects of some **Machine Learning** techniques in Artificial Intelligence to enable you to solve practical problems using these learning techniques.

0.2 What this module is about

Purpose

Machine Learning (COS4852) is will expose you to the theory and inner working of various machine learning algorithms, so that you are able to understand and implement these algorithms. The field of Machine Learning has grown to be quite large and in this module you will see some the basic building blocks behind the latest algorithms.

Pre-requisites

You should have a solid background in theoretical computer science, including a 3rd year module in Artificial Intelligence, as well a solid working knowledge of linear algebra, vector and matrix manipulation as well as basic probability theory.

Syllabus

The following topics of Machine Learning are covered in COS4852:

- The theoretical background to learning theory in the form of Concept Learning:
    - General-to-specific ordering
    - Find-S
    - Version spaces
    - Candidate-Elimination
    - List-Then-Eliminate
    - Bias in Learning
- Instance-based learning – may include the following:
    - k-Nearest Neighbours
    - IB1, IB2, IB3
    - Kernel machines
    - kd-trees
    - Case-based reasoning
- Decision Trees
    - ID3
- Neural Networks
    - Perceptrons
    - Backpropagation
    - Deep-Learning
- Bayesian Learning
    - Bayes' Theorem
    - MAP
- Genetic Algorithms
    - Genetic Programming

Outcomes

If you master the material, by the end of this module you should be able to do the following:

- Discuss the theoretical background to learning using concept learning.
- Implement and solve problems using algorithms related to concept learning, such as Find-S and Find-G.
- Implement and solve problems using algorithms related to instance-based learning, such as kNN and IB1.
- Implement and solve problems using ID3.
- Implement and solve problems using Perceptrons and the Backpropagation algorihms.
- Discuss Deep-Learning algorithms.
- Discuss Bayes' Theorem.
- Implement and solve problems using MAP.
- Implement and solve problems using Genetic Algorithms.

0.3 Resources

myUnisa

COS4852 will be run as an online-only module, in other words all the material and information you will need for this module will be available on *myUnisa*. More information about exactly this works will be available on the module site.

School of Computing website - OSPREY

The School of Computing has a webserver called Osprey on which information about the school is located. The URL for this server is http://osprey.unisa.ac.za/.

The above mentioned page also contains a link (called **OSS repository**) where you can find a repository of open-source software, from which you can download software. You may find this useful for the assignments for COS4852.

Prescribed book

There is no prescribed book for COS4852. The study material on myUnisa will point you to sources on the internet, and we will make some material available via the ***Additional Resources*** tool on *myUnisa*.

Programming and Software

This module will require you to develop code to get familiar with the practical aspects of the machine learning algorithms. You will need to get familiar with Python. The ideal environment for playing with the code is a Jyputer notebook, which is a browser-based environment for coding and writing notes.

0.4 Overview of the COS4852 site on *myUnisa*

There are a number of tools that have been activated for COS4852 on this *myUnisa* site. The most important ones are:

- *Lessons***:** These contain comments on the relevant sections of the prescribed book, activities you should do in preparation for the assignments, as well as the assignment questions.
- *Additional resources***:** Extra documents referred to in the lessons are placed here.
- *Assignments:* This is where you can submit assignments, and receive your marked assignments back.
- *Announcements:* From time to time, your lecturers will post important messages for the attention of all COS4852 students.
- *Discussion forum:* Here you can participate in online discussions with your fellow students.

0.5 Do this first

Here are the things you need to do in preparation for your COS4852 studies:

Activate your myLife email

Please activate your myLife email account (if you haven't done so yet). If you don't like the myUnisa interface for sending and receiving myLife emails, you can redirect your myLife emails to your personal email address. If you check your email regularly, you will receive any messages or announcements as soon as they are sent.

Internet access

Since this module is run online, and uses resources from the internet, you will have to have access to the internet.

Software

You will have to write code for the assignments in this module. Make sure that you have the Jupyter notebook installed.

0.6 Assessment

Assessment for COS4852 consists of *three assignments* (which represent the formative assessment opportunities) and a *written examination* (which represents the summative assessment).

See the *Assignments* tool for the due dates for the assignments and the portfolio.

Suggested study plan

We recommend that you work through the material in each lesson at least two weeks before the due date of each assignment, to give you enough time to complete and submit the assignment.

After submitting the last assignment, you should start working on the portfolio, as will comprise a substantial volume of work.

Assignments

All three assignments are compulsory, and contribute to your year mark.

The tasks for the assignments will be given in the study units, not necessarily one assignment per unit.

All the assignments are intended to be a learning exercise. We therefore advise you to put in a serious effort to complete the assignment to the best of your ability. Each assignment also contributes a significant portion of your final mark. Plagiarism and copying from each other will therefore be penalised severely.

All assignments must be submitted via myUnisa. Assignments that are emailed, or that are submitted as hard-copy will not be marked.

Make sure you follow the procedures outlined in the Studies@Unisa brochure when submitting your assignments. Assignments should be submitted by the due dates. Please do not contact us for extensions.

Preparing and submitting assignments

For each assignment where you need to write code, you must submit a single report in PDF format (with an extension **.pdf**).

Discussion of assignments

A discussion of the each assignments will be posted after the due date of the assignment.

Examination

There is a formal examination for this module.

Final mark calculation

Your year mark is calculated as a weighted sum of the three assignments, as follows:

$$YM = (A_1 \times 0.3) + (A_2 \times 0.3) + (A_3 \times 0.4)$$

where A_1, A_2, and A_3 are your marks for the assignments.

Your final mark (*FM*) is calculated as as a weighted sum of your yearmark (YM) and your examination mark (*EM), as* follows:

$$FM = (YM \times 0.2) + (EM \times 0.8)$$

In other words, your year mark contributes 20% and your exam mark 80% to your final mark.