## Machine Learning courses – Long version

## (CORE I) General Machine Learning & Data Science – 26 hours

**(4 hours) An overview of modern day Data Analytics and Machine Learning**

In this module, we will explain how modern day machine learning and AI technologies have transformed various industries. We do so by showing a range of demos and examples of the relevant techniques developed for companies in each of the specialised. Note that some of the example settings may change depends on whom the attendees are

**[exercise]**

**There is no practical associated with this module. Instead, there will be a few interesting publically available references we would like participants to read/watch**

**(4 hours) Machine Learning (ML) fundamentals**

In this module, we also explain some of the fundamental knowledge in machine learning, such as what is supervised vs unsupervised learning; what are examples of regression and classification problems. What is variance-biased trade-offs; how to measure classification accuracy in terms of precision, recall and F1 scores, and AUC. Dimensionality Reduction techniques such as Principal Component Analysis and modern t-SNE will also be presented.

We will also introduce attendees **gently to Python programming, including numpy and scikit-learn libraries**

**[exercise]**

1. **Exercise involving basic Python programming, and using numpy, scikit-learn and pandas library**
2. **Read a *synthetic* information data using Python (pandas library); compute simple statistics from it**
3. **Apply a simple dimension reduction algorithm, such as PCA on the project information dataset**
4. **apply a simple binary classification to predict profit (1)/loss (0) on the project information dataset**

**(5 hours) Cluster analysis**

In this module, we explain how classical clustering algorithms such as K-mean, Gaussian Mixture Model (GMM), Affinity Propagation (AP) work. We also include algorithm dealing with clustering when data are partially missing some of its attributes. Some techniques to penalise the excessive number of clusters, such as BIC and AIC will also be presented

Optimization techniques such as expectation-maximization used in cluster analysis will be illustrated here from a high-level

**[exercise]**

1. **using Python to program K-means and GMM algorithm with and without using scikit-learn function**
2. **to apply several clustering algorithms on synthetic data**
3. **visualise the dataset clusters using python matplotlib**

**(5 hours) Classification and Regression**

We explain some of the classical regression techniques including Linear, Polynomial in the Least Square sense. We will also explain what Mixed Effects models are.

We also present some classification algorithm bear the name of regression: namely, logistic and multinomial regression.

**[exercise]**

1. **using Python to perform regression model on a synthetic dataset to predict its project cost; this is to performed using:**
2. **linear regression,**
3. **polynomial regression and**
4. **linear regression with mixed effects model**
5. **using Python to perform logistic regression on a synthetic project dataset to predict its profitability (1=profit, 0=loss)**

**(4 hours) Time series modelling, and** [**Continuous and Discrete State Dynamic Systems**](http://www-staff.it.uts.edu.au/~ydxu/ml_course/dynamic_model.pdf)

We show the general concepts of sequential dataset, to distinguish between set and series. We show classical forecast models such as Autoregressive integrated moving average (ARIMA) and the idea of seasonal effects. We also present other continuous and discrete state systems such as Kalman Filter and Hidden Markov Model, the concept of states, observations and transitions probabilities between consecutive states

**[exercise]**

1. **Using python to program a simple inventory modelling and prediction using Autoregressive integrated moving average (ARIMA)**
2. **Using python to program a simple inventory modelling and prediction using Hidden Markov Model (HMM)**

**(4 hours) Recommendation System**

We explain what is collaborative Filtering, Non-Negative Matrix Factorization, Alternating Lease Square, restricted Boltzmann machine approach and Factorization Machines to Recommendation system; We will also illustrate the winning solution of Netflix Challenge in Recommendation System.

**[exercise]**

**simple exercise on** recommendation system using:

1. **Collaborative filtering**
2. Non-Negative Matrix Factorization;