FEM21045-20 Machine Learning in Finance

Block 1 (Sep-Oct 2020)

COURSE SYLLABUS

Faculty

Coordinator

Dick van Dijk (djvandijk@ese.eur.nl)

Instructors

Dick van Dijk [**DvD**] (djvandijk@ese.eur.nl)

Ilker Birbil [**IB**] (birbil@ese.eur.nl)

Utku Karaca [**UK**] (karaca@ese.eur.nl)

Karel de Wit [**KdW**] (dewit@ese.eur.nl)

Purpose & Learning goals

This course aims to provide an introduction to machine learning techniques, focusing on those methods that are useful and popular in financial applications. Specific learning goals include:

- Learn what machine learning is
- Study the most important prediction (regression/classification) techniques and clustering methods
- Be able to implement and apply these techniques
- Gain insight into the theory behind some of the most frequently used methods

The course has substantial overlap with the course FEM31002-20 Machine Learning. The main difference concerns the additional lectures discussing empirical applications of machine learning techniques in finance and (macro)economics.

Content

The main topics covered in this course are:

- Regularization
- Classification and Regression Trees, Forests and Ensemble Methods
- Support Vector Machines
- Clustering
- Neural Networks

Course outline

- Week 1: Introduction, Resampling and Model Evaluation
- Week 2: Regularization
- Week 3: Trees, Forests and Ensembles
- Week 4: Unsupervised Learning
- Week 5: Support Vector Machines
- Week 6: Neural Networks
- Week 7: Reinforcement Learning

See the Appendix for a more detailed course outline and schedule.

Lectures, tutorials, and Q&A sessions

Due to the COVID-19 related restrictions and limitations for on-campus teaching, lectures are provided online. For each week, links to pre-recorded lectures covering the relevant topic will be provided via Canvas. The lectures will cover theory and empirical applications; either combined/integrated in a single lecture, or in separate (short) lectures. Slides for the lectures will also be made available on Canvas.

For each topic, a live Q&A session will be organized in Zoom on Wednesdays, from 16:00-17:45 in week 1 and from 10:00-11:45 in weeks 2-7. Questions for these Q&A sessions can be submitted via the Discussion forum on Canvas, see the links in the respective Modules. Each forum will be kept open until a few days before the live lecture or tutorial. During the live sessions, the instructors will first go through all the questions compiled from the forums, and then if time permits, they will continue with the other questions from the class.

For the assignments, a pencast tutorial will be provided on Canvas.

For each assignment, a live Q&A session will be organized in Zoom on Fridays, from 11:00-11:45 in weeks 2-7. A discussion forum for questions about the assignments will be available as well, with the same set-up as described above for the Q&A sessions related to the lectures.

Course material / Literature

The course is based on the textbook

> Hastie, T., R. Tibshirani, and J. Friedman (2009), *The Elements of Statistical Learning - Data Mining, Inference, and Prediction*, 2nd Edition, Springer (available online here)

For each topic, the relevant sections of the book will be indicated on Canvas [both compulsory and further reading]

For the different topics covered during the course, additional material in the form of recent journal articles and working papers will be provided. Some of these papers you are required to read. For those who want to dig deeper into a certain topic, we will also provide some supplementary material. Note that reading this supplementary material is optional and you will not be tested from that content.

Software

We will use Python for (part of) the assignments. The website of the textbook also includes data sets, codes as well as R functions. We recommend installing the Anaconda Distribution for Python.

Assignments

The course involves six assignments. Assignments 3 and 5 are mandatory, and contribute (20% each) to the final course grade. These two assignments should be made in groups of three students. The other (voluntary) assignments (1, 2, 4 and 6) should be made individually.

Assessment and grading

The course assessment consists of two components (with their weights in the final course grade given in parentheses):

I. Written exam [closed-book; open questions] (60%)
II. Two assignments (40% in total [20% each]) – see above for details

In order to pass the course, the final course grade has to be 5,5 or higher (on the usual 10-point grading scale) There are no minimum grade requirements for any of the assessment components.

Course Policies

Behaviour and Communication

Arrive in time for the class. Please refrain from using computers and phones for anything but activities related to the class. Voice recording of the lectures are strictly prohibited. Eating and drinking are allowed in class but please make sure it does not affect the lecture. When communicating with the lecturer and fellow students using online platforms, be aware of the communication etiquette. Information about code of conduct at Erasmus University Rotterdam is available at: https://my.eur.nl/en/ese/master/code-conduct.

Academic Integrity

In agreement with Erasmus University Rotterdam integrity code, you are expected to maintain an attitude and professionalism, teamwork and fair play: https://www.eur.nl/en/about-eur/strategy-and-policy/integrity. Treat your fellow students and the lecturer with respect to enable a pleasant and stimulating work and study environment.

Academic Honesty

Students are required to comply with the university policy on academic honesty. To understand what constitutes cheating and plagiarism, see https://www.eur.nl/en/about-eur/vision/cheating-and-plagiarism.

Copyright

Erasmus University Rotterdam holds the copyrights to the teaching material for this course. Therefore, distribution of the course material to a third party for commercial purposes constitutes an infringement of the copyright. More information is available at https://www.eur.nl/en/copyright-information-point.

Appendix: Course outline in more detail

Week 1: Introduction [IB]

- Overview of the course
- Supervised Learning vs. Unsupervised Learning
- Train-test errors and overfitting
- Bias vs. Variance
- Bayes Classifier vs. K-Nearest Neighbor (KNN)
- Cross validation and bootstrap
- Model evaluation and algorithm comparison

Week 2: Regularization [IB]

- Shrinkage: Ridge Regression and Lasso
- Least Angle Regression
- Elastic Net
- Integer Programming models

Week 3: Trees, Forests and Ensembles [IB]

- Decision Trees
- Random Forest
- Boosting
- Interpretability

Week 4: Unsupervised Learning [DvD]

- K-means, K-medoids
- Hierarchical Clustering
- Spectral Clustering
- Nonnegative Matrix Factorization
- Page Rank Algorithm

Week 5: Support Vector Machines [IB]

- Geometry of support vector machines
- Dual problem
- Path to kernels
- Interpretable kernels

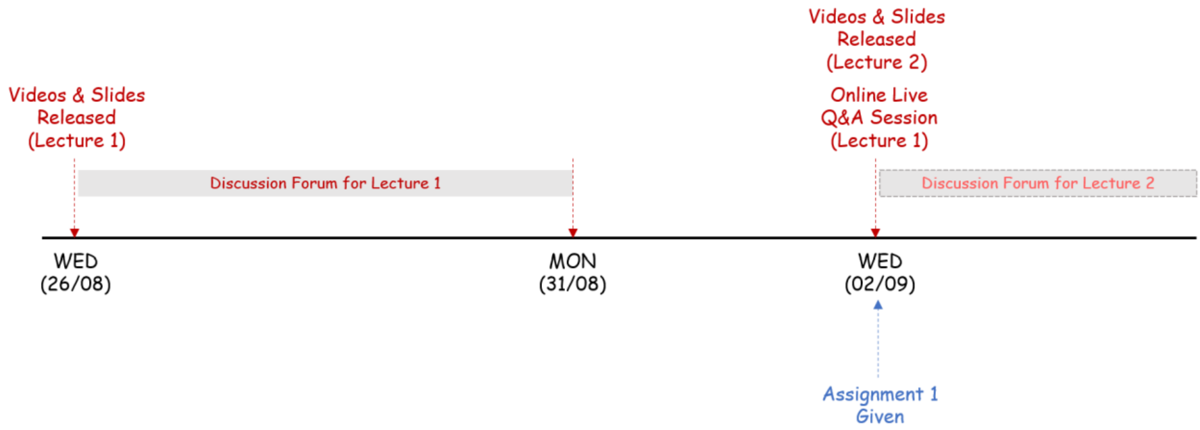
Week 6: Neural Networks and Deep Learning [DvD]

- Topology of artificial neural networks
- Backpropagation
- Vanishing gradient
- Convolutional Neural Networks
- Different deep learning structures

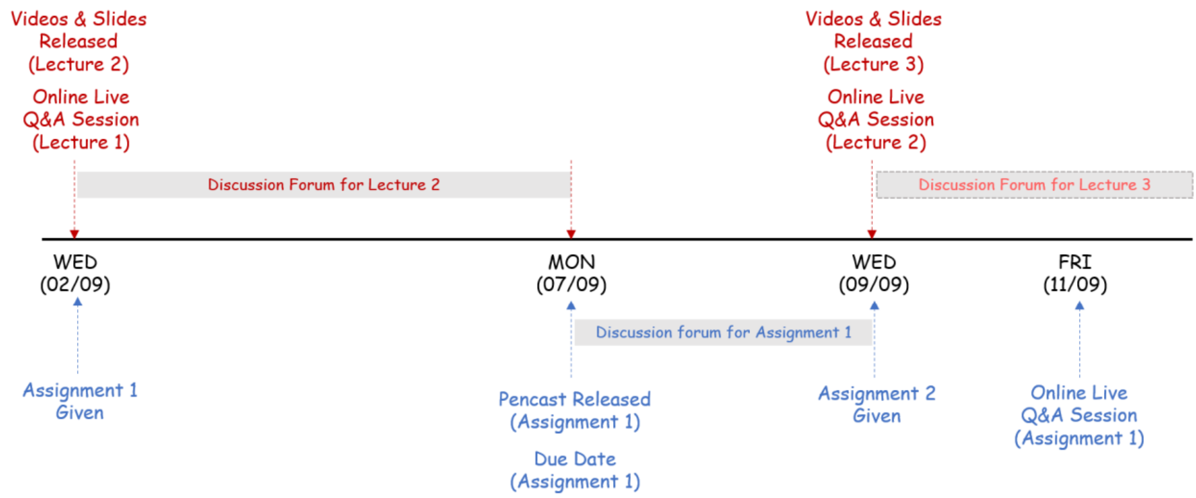
Week 7: Reinforcement Learning [UK]

- States, actions, rewards
- K-armed Bandit
- Model-based vs. Model-free
- Learning algorithms
- Relation to Markov Decision Processes

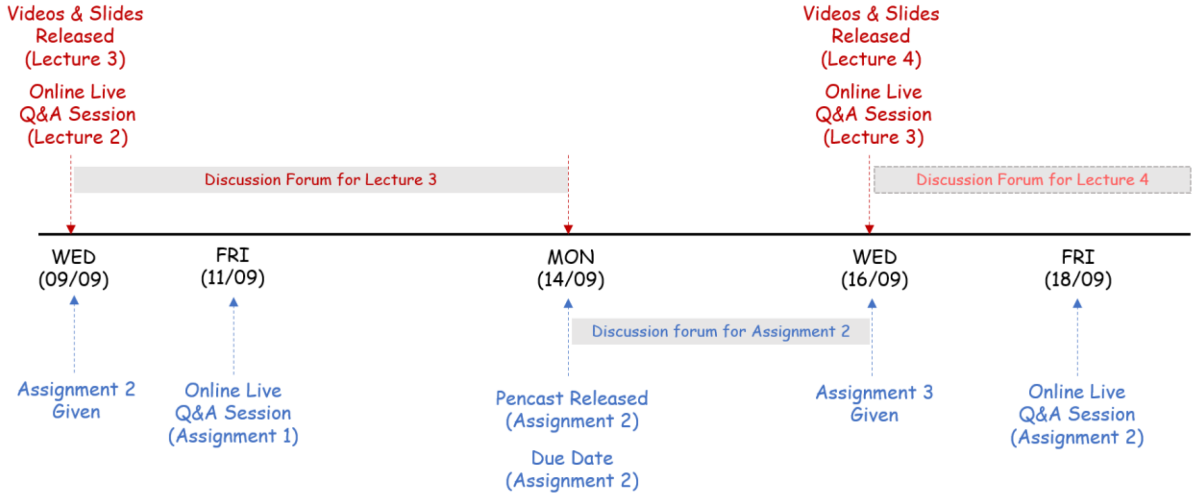
WEEK 0-1



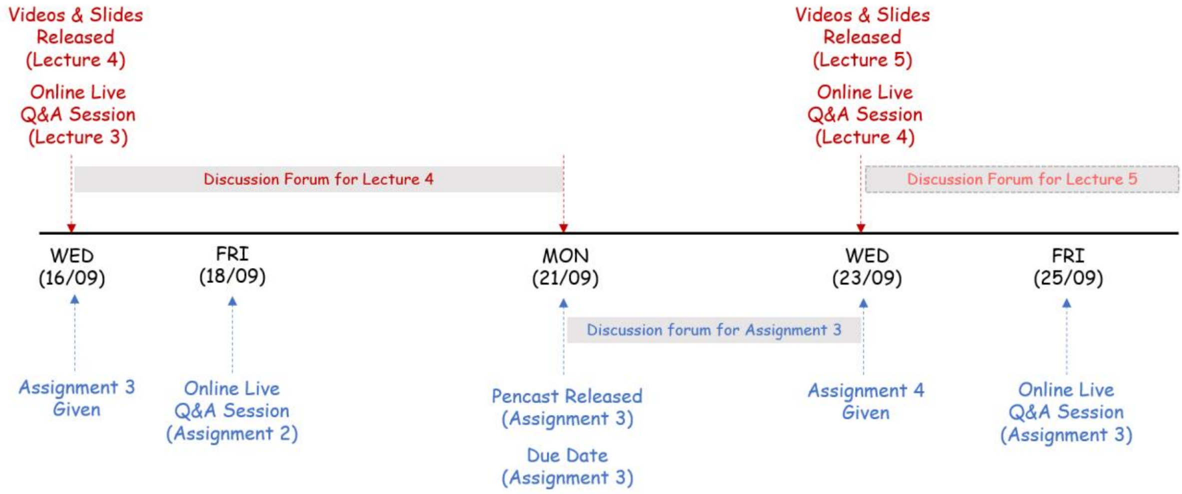
WEEK 1-2



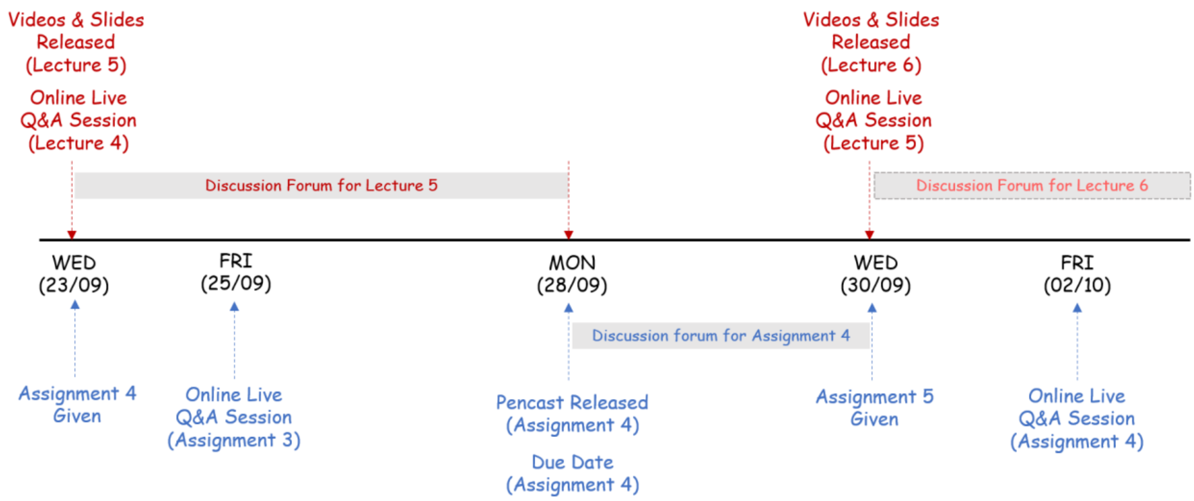
WEEK 2-3



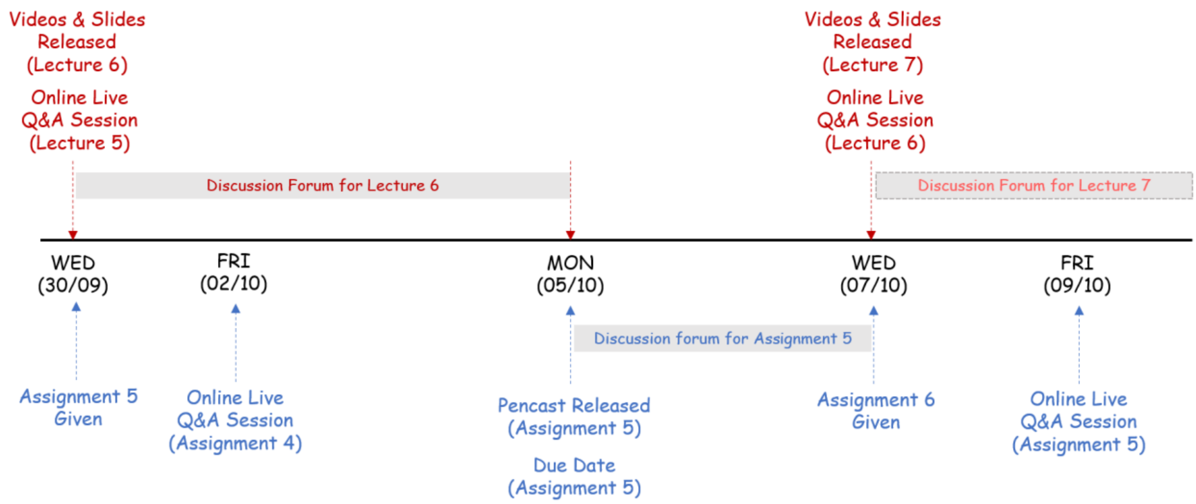
WEEK 3-4



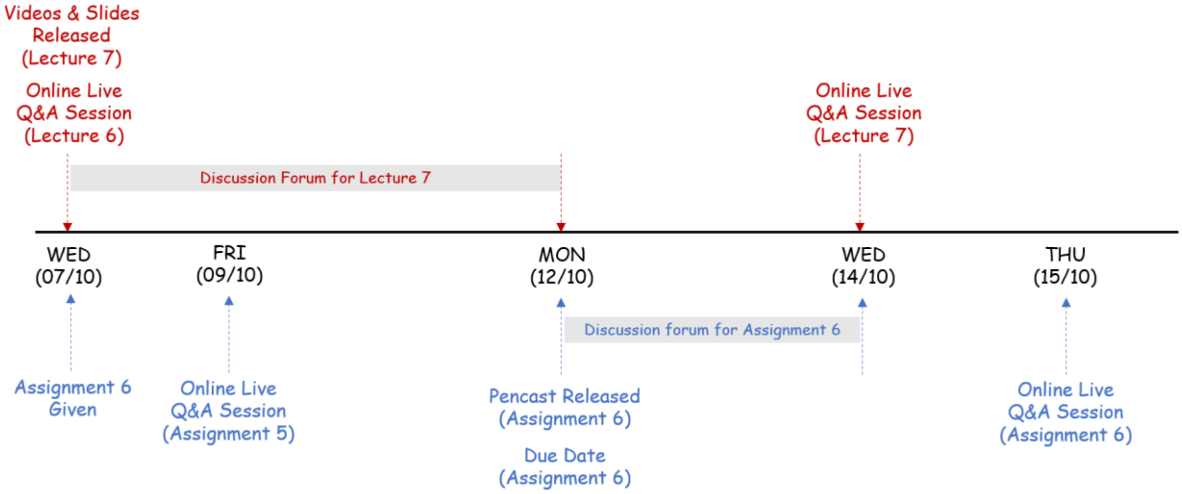
WEEK 4-5



WEEK 5-6



WEEK 6-7


Videos & Slides Released (Lecture 7)
Online Live Q&A Session (Lecture 6)
Online Live Q&A Session (Lecture 7)
Discussion Forum for Lecture 7
WED (07/10)
FRI (09/10)
MON (12/10)
WED (14/10)
THU (15/10)
Discussion forum for Assignment 6
Assignment 6 Given
Online Live Q&A Session (Assignment 5)
Pencast Released (Assignment 6)
Due Date (Assignment 6)
Online Live Q&A Session (Assignment 6)