COURSE: Supervised Learning

INSTRUCTORS: Gianluca Cubadda (Ph.D. 1994); Antonio Parisi (Ph.D. 2008)
EMAIL: gianluca.cubadda@uniroma2.it
    antonio.parisi@uniroma2.it
WEB PAGE: http://directory.uniroma2.it/index.php/chart/dettagliDocente/3916
    http://directory.uniroma2.it/index.php/chart/dettagliDocente/8215

COURSE DESCRIPTION
The course provides an introduction to supervised learning, focusing on both regression and classification problems. Empirical applications will be illustrated using updated software tools.

LEARNING OUTCOMES
- ✓ Understand regression models.
- ✓ Understand supervised classification methods.
- ✓ Gain practical experience in predicting a target variable using many predictors.

METHODOLOGY
Theoretical lessons and practice using R and Matlab.

ASSESSMENT
Written exam; weighting: 80%
Project; weighting: 20%

OUTLINE
- ➢ Introduction to Statistical Learning
    1. Overview
    2. Statistical Learning and Regression Models
    3. The Dimensionality Problem
    4. Model Accuracy and the Bias-Variance Trade-Off
- ➢ The Linear Regression Model
    1. Interpretation and Estimation of Coefficients
    2. Prediction
    3. Model Selection
    4. Qualitative Predictors
    5. Interaction Terms
    6. Non-linear Transformations of the Predictors
    7. Outliers and Influential Observations
- ➢ Resampling Methods
    1. Cross-validation
    2. The Bootstrap
- ➢ Model Selection and Regularization Methods
    1. Subset Selection
    2. Shrinkage (Ridge e LASSO)
    3. Dimension Reduction (Principal Component Regression and Partial Least Squares)

- ➢ Tree-Based Methods for Regression Problems
    1. Regression Trees
    2. Bagging
    3. Random Forests
    4. Boosting
- ➢ Moving Beyond Linearity
    1. Polynomial Regression
    2. Step Functions
    3. Basis Functions
    4. Regression Splines
    5. Smoothing Splines
    6. Local Regression
    7. Generalized Additive Models
- ➢ Classification
    1. Overview
    2. Logistic Regression
    3. Linear Discriminant Analysis
    4. Comparison of the Methods
- ➢ Tree-Based Methods for Classification Problems
    1. Classification Trees
    2. Bagging Classifier
    3. Random Forest Classifier
    4. Boosting Classifier
- ➢ Support Vector Machines
    1. Maximal Margin Classifier
    2. Support Vector Classifiers
    3. Support Vector Machines
    4. SVMs with More than Two Classes
    5. Relationship to Logistic Regression

TEXTBOOKS

Hastie T., Tibshirani R., and J. Friedman (2011), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed., Springer: New York.
   Web page: https://web.stanford.edu/~hastie/ElemStatLearn/
Gareth J., Witten D, Hastie T., and R. Tibshirani (2013), *An Introduction to Statistical Learning: With Applications in R*, Springer: New York.
   Web page: https://www-bcf.usc.edu/~gareth/ISL/

ADDITIONAL SUGGESTED READING

Izenman A.J, (2013) *Modern Multivariate Statistical Techniques Regression, Classification, and Manifold Learning*, 2nd ed., Springer: New York.