

## **CORSO: Supervised Learning**

DOCENTI: Gianluca Cubadda (Ph.D., 1994); Giovanni Trovato (Ph.D., 2001)

EMAIL: [gianluca.cubadda@uniroma2.it](mailto:gianluca.cubadda@uniroma2.it)

[giovanni.trovato@uniroma2.it](mailto:giovanni.trovato@uniroma2.it)

PAGINE WEB: <http://directory.uniroma2.it/index.php/chart/dettagliDocente/3916>

<http://directory.uniroma2.it/index.php/chart/dettagliDocente/5228>

### **DESCRIZIONE DEL CORSO**

Il corso fornisce un'introduzione ai metodi di apprendimento supervisionato, focalizzandosi sia su problemi di regressione che di classificazione.

### **OBIETTIVI DI APPRENDIMENTO**

- ✓ Comprendere il modello di regressione.
- ✓ Comprendere i metodi di classificazione supervisionati.
- ✓ Acquisire abilità pratica nel predire una variabile obiettivo mediante l'uso di molti predittori.

### **METODOLOGIA**

Lezioni teoriche ed esercitazioni, per le quali si farà l'uso del linguaggio R.

### **VALUTAZIONE**

Esame scritto; peso: 80%

Progetto di gruppo; peso: 20%

### **PROGRAMMA**

- 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
- 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

### **LIBRI DI TESTO**

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/>

### **LETTURE SUGGERITE**

Hastie T., Tibshirani R., and M. Wainwright (2015) *Statistical Learning with Sparsity. The Lasso and Generalizations*, Chapman and Hall: New York.

Web page: <https://hastie.su.domains/StatLearnSparsity/>

Efron, B., and T. Hastie (2016) *Computer Age Statistical Inference: Algorithms, Evidence and Data Science*, CUP: New York.

Web page: <https://hastie.su.domains/CASI/>