UNIVERSITÀ DI PAVIA

Anno Accademico 2020/2021

STATISTICAL LEARNING THEORY

Anno immatricolazione	2019/2020
Anno offerta	2020/2021
Normativa	DM270
SSD	ING-INF/04 (AUTOMATICA)
Dipartimento	DIPARTIMENTO DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE
Corso di studio	COMPUTER ENGINEERING
Curriculum	Computer Science and Multimedia
Anno di corso	2°
Periodo didattico	Primo Semestre (28/09/2020 - 22/01/2021)
Crediti	6
Ore	45 ore di attività frontale
Lingua insegnamento	English
Tipo esame	SCRITTO
Docente	DE NICOLAO GIUSEPPE (titolare) - 2 CFU DE NICOLAO GIUSEPPE (titolare) - 4 CFU
Prerequisiti	Matrix algebra; elements of probability: scalar and vector random variables; elements of statistics: estimators and their properties.
Obiettivi formativi	Knowledge of main learning methods for classification and regression, of their properties and limitations. Ability to translate an experimental learning problem into a statistical formulation and select an appropriate method for its solution.
Programma e contenuti	Introduction: Supervised and Unsupervised Learning. Statistical Learning: Statistical Learning and Regression, Curse of Dimensionality and Parametric Models, Assessing Model Accuracy and Bias-Variance Trade-off, Classification Problems and K-Nearest Neighbors. Linear Regression: Simple Linear Regression and Confidence Intervals,

	Hypothesis Testing, Multiple Linear Regression, Model Selection, Interactions and Nonlinearity. Classification: Introduction to Classification, Logistic Regression and Maximum Likelihood, Linear Discriminant Analysis and Bayes Theorem, Naive Bayes. Resampling Methods: Estimating Prediction Error and Validation Set Approach, K-fold Cross-Validation, Cross-Validation: The Right and Wrong Ways, The Bootstrap. Linear Model Selection and Regularization: Linear Model Selection and Best Subset Selection, Stepwise Selection, Estimating Test Error Using Mallow's Cp, AIC, BIC, Adjusted R-squared, Cross-Validation, Shrinkage Methods and Ridge Regression, The Lasso, Principal Components Regression and Partial Least Squares. Moving Beyond Linearity: Polynomial Regression, Piecewise Polynomials and Splines, Smoothing Splines, Local Regression and Generalized Additive Models. Tree-Based Methods: Decision Trees, Classification Trees and Comparison with Linear Models, Bootstrap Aggregation (Bagging) and Random Forests, Boosting. Support Vector Machines: Support Vector Classifier, Kernels and Support Vector Machines. Unsupervised Learning: Unsupervised Learning and Principal Components Analysis, K-means Clustering. The fallacies of learning: regression to mediocrity, the covariate shift, statistical significance vs practical significance, correlation is not causation, observational vs experimental studies.
Metodi didattici	Lectures, practical class.
Testi di riferimento	Friedman, J., Hastie, T., & Tibshirani, R. (2001). The elements of statistical learning (Vol. 1, No. 10). New York: Springer series in statistics.
Modalità verifica apprendimento	Written examination: two theory-based questions and two practical ones.
Altre informazioni	Written examination: two theory-based questions and two practical ones.
Obiettivi Agenda 2030 per lo sviluppo sostenibile	$lbl_legenda_sviluppo_sostenibile