

Data Analytics, a.a. 2020-2021

Argomenti del Corso

Docenti: Prof. Giuseppe Lisanti, Prof. Marco Di Felice
{giuseppe.lisanti, marco.difelice3}@unibo.it

December 11, 2020

1. Data Processing, Plotting and Analytics con Python

- Libreria Pandas
- Libreria NumPy
- Libreria Matplotlib
- Libreria Seaborn
- Framework Scikit-Learn
- Framework PyTorch

2. Data Fundamentals

- Data, Dataset, Attributo/Variabile, Variabile categoriche/ordinale/etc
- Statistica descrittiva: media, mediana, moda, varianza, quartili
- Apprendimento supervisionato vs non supervisionato

3. Data Acquisition

- Acquisizione Flat vs Stream
- Acquisizione da file ed encoding delle variabili categoriche

4. Data Visualization

- Fundamentals: Marker, Channel, Color Encoding
- Metodi per la visualizzazione di variabili singoli (categoriali/numerici) vs variabili multiple

5. Data Pre-Processing

- Data Cleaning: gestione valori mancanti e outlier
- Normalizzazione, Standardizzazione e Scaling delle variabili
- Riduzione della dimensionalità: PCA, LDA

- Metodi di shrinkage: Ridge regression, Lasso
6. **Regressione Lineare**
 - Regressione Lineare Semplice
 - Regressione Lineare Multipla
 - Logistic Regression
 7. **Metodi di Classificazione basati su region splitting**
 - K-Nearest Neighbour (KNN)
 - Maximal Margin Hyperplane
 - Support Vector Machines (SVM) e funzioni kernel
 8. **Metodi Bayesiani**
 - Formula di Bayes ed Optimal Bayesian Classifier
 - Naive Bayes
 - Quadratic Discriminant Analysis (QDA)
 9. **Metodi basati su Alberi Decisionali**
 - Alberi di classificazione vs Alberi di regressione
 - Metriche di purezza per la costruzione dell'albero
 - Metodi ensemble: Bagging, Random Forest, Boosting
 10. **Feedforward Neural Networks (NN)**
 - Funzione di costo
 - Gradient Descent e tecniche di back-propagation
 - Regularizzazione, ottimizzazione e training
 11. **Recurrent Neural Networks (RNN)**
 - Definizione
 - Encoding e decoding
 - Architettura di Long Short Term Memory Networks (LSTM)
 12. **Convolutional Neural Networks (CNN)**
 - Definizione di Convoluzione
 - Padding & Striding
 - Livelli ed Architetture di CNN
 13. **Unsupervised Learning**
 - Distance-based Clustering: K-Means
 - Gaussian Mixture Models (cenni)
 - Density-based Clustering: DBSCAN
 - Hierarchical clustering