

Contextualization and contextualization

Víctor Mario Noble Ramos, MSc.

2024-08-29

Outline for section 1

1 Course presentation

2 Let us start: Fundamental concepts

Course presentation

Welcome to introduction to Machine Learning: algorithms and theory!

Presentation

Victor Mario Noble Ramos

- Industrial engineer with a master in Production Engineering from *Universidade Federal de São Carlos - UFSCar*.
- +10 years of experience in teaching and academic research.
- Graduated with honors: DS4A - Correlation One & MINTIC
- Research topics: Operations research, Data Science, Engineering education.

Presentation

Victor Mario Noble Ramos

- Industrial engineer with a master in Production Engineering from *Universidade Federal de São Carlos - UFSCar*.
- +10 years of experience in teaching and academic research.
- Graduated with honors: DS4A - Correlation One & MINTIC
- Research topics: Operations research, Data Science, Engineering education.

What about you?

Course content

Course objective:

Goal

To provide students with a solid understanding of the fundamental principles of machine learning, from data preparation and cleaning to the implementation and evaluation of basic models.

Course content

Course objective:

Goal

To provide students with a solid understanding of the fundamental principles of machine learning, from data preparation and cleaning to the implementation and evaluation of basic models.

Macrounits

- 1 Generalities of the field
- 2 Data preparation, EDA and visualization
- 3 ML models
- 4 Common challenges

1. Generalities of the field

Goal

To know and explore different concepts, history and the context of the Data Sciences and the Machine Learning field, as well as to refresh previous notions of the discipline.

1. Generalities of the field

Goal

To know and explore different concepts, history and the context of the Data Sciences and the Machine Learning field, as well as to refresh previous notions of the discipline.

Contents 1/3

- Concepts: Machine learning, Statistical learning, Deep learning, Artificial intelligence, Data science, etc.
- Importance and history of the field.
- Applications

1. Generalities of the field

Goal

To know and explore different concepts, history and the context of the Data Sciences and the Machine Learning field, as well as to refresh previous notions of the discipline.

Contents 2/3

- Refresher of statistics
 - Random variables and data types
 - Descriptive statistics
 - Probability distributions

1. Generalities of the field

Goal

To know and explore different concepts, history and the context of the Data Sciences and the Machine Learning field, as well as to refresh previous notions of the discipline.

Contents 3/3

- Refresher of programming
 - Data structures: the dataframe, series
 - Control flow statements
 - Work with libraries (modules)

2. Data preparation, EDA and visualization

Goal

To teach to the students how to prepare and clean the data as well as how to explore and understand the data before applying ML models.

2. Data preparation, EDA and visualization

Goal

To teach to the students how to prepare and clean the data as well as how to explore and understand the data before applying ML models.

Contents - 2.1. Data preparation

- Introduction to pandas.
- Data importing, storage and file formats.
- Handling missing values, error detection, and correction.
- Data normalization and standardization.
- Feature engineering: creation and selection of features.

2. Data preparation, EDA and visualization

Goal

To teach to the students how to prepare and clean the data as well as how to explore and understand the data before applying ML models.

Contents - 2.2. Data wrangling

- Importing and merging data from different tables [and query language between tables (Optional)].
- Reshaping and pivoting.
- String manipulation.

2. Data preparation, EDA and visualization

Goal

To teach to the students how to prepare and clean the data as well as how to explore and understand the data before applying ML models.

Contents - 2.3. EDA and visualization

- Plotting with matplotlib
- Other libraries for plotting
- Descriptive univariate statistics and probability distributions.
- Outliers identification and handling.
- Bivariate and multivariate relations (Correlation matrix)
- Dimensionality reduction.
- Preliminary hypotheses testing.

3. ML models

Goal

To comprehend and apply some of the ML/ST models used for supervised learning in real and synthetic data.

Contents - 3.1. Supervised learning 1/3

- Regression and classification problems
 - Overview and refresher of linear regression
 - Overview and refresher of logistic regression
 - Linear model selection and regularization

3. ML models

Goal

To comprehend and apply some of the ML/ST models used for supervised learning in real and synthetic data.

Contents - 3.1. Supervised learning 2/3

- Extensions of linear-based methods
 - Polynomial regression
 - Step functions
 - Regression splines
 - Smoothing splines

3. ML models

Goal

To comprehend and apply some of the ML/ST models used for supervised learning in real and synthetic data.

Contents - 3.1. Supervised learning 3/3

- Tree-based methods
- Support vector machines
- Introduction to deep learning

3. ML models

Goal

To comprehend and apply some of the ML/ST models used for supervised learning in real and synthetic data.

Contents - 3.2. Unsupervised learning

- Fundamentals and problem statement
- Techniques
 - K-means clustering
 - Hierarchical clustering

3. Common challenges and final remarks

Goal

To comprehend and apply some of the ML/ST models used for supervised learning in real and synthetic data.

Contents

- High dimensional data
- Unbalanced data
- Unlabeled data
- Overfitting
- Closing remarks

Evaluation

The course will be evaluated by three types of activities:

- Homework socialization (40%)
- Midterm exam (30%)
- Final exam (30%)

References

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). ***An Introduction to Statistical Learning (2nd ed.)***. Springer.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer New York.
<https://doi.org/10.1007/978-0-387-84858-7>
- Géron, A. (2023). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (3rd ed.)*. O'Reilly Media Inc.
- McKinney, W. (2021). ***Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython (3rd ed.)***. O'Reilly Media.
- Vicente Cestero, E., & Mateos Caballero, A. (2023). *Inteligencia Artificial: Fundamentos matemáticos, algorítmicos y metodológicos*. España: Eloy Vicente Cestero. ISBN 978-84-09-46911-6.
- Berzal, F. (2018). *Redes Neuronales & Deep Learning*. España. Fernando Berzal. ISBN 1-7312-6538-7

Outline for section 2

1 Course presentation

2 Let us start: Fundamental concepts

Supervised and unsupervised learning

Unsupervised learning