

# MACHINE LEARNING per la FISICA

**Corso seminariale**

**Secondo semestre**

**Per le lauree:** Triennale e Magistrale

**Crediti formativi:** 3 CFU

**Titolo:** Machine Learning per la Fisica

**Docente:** Pierluigi Bortignon

## **Prerequisiti**

Basi di statistica e programmazione, indipendente dal linguaggio. Aver completato il corso di Laboratorio 1 e Informatica aiuta molto la comprensione.

## **Descrizione**

Il machine learning ha dimostrato grandi potenzialità nel migliorare il rendimento di molti processi in vari campi.

Machine learning sono algoritmi computazionali in grado di classificare informazioni e predire accuratamente esiti di processi, senza essere programmate esplicitamente per farlo, imparando e sfruttando correlazioni di dati in ingresso.

Grazie all'incremento della quantità di dati disponibili e maggiore potenza di calcolo grazie a HPC si stanno sviluppando sempre più sofisticate tecniche di machine learning con risultati sorprendenti.

In fisica l'utilizzo di algoritmi di machine learning ha portato a scoperte importanti prima di quanto ci si aspettasse ed è sempre più presente in molteplici aree di ricerca, dall'analisi dei dati, la progettazione di rivelatori, la rivelazione di anomalie, alla generazione di simulazioni.

## **Programma del corso**

### *Concetti trattati*

Regression. Classification. Supervised learning. Unsupervised learning. Reinforcement learning. Function approximation, Model, Hyper-parameters, Generalization, Regularization. Decision trees (bagging, boosting, random forest). Artificial neural network, the Multi Layer Perceptron, Gradient descent techniques. Deep networks. Convolutional networks. Recursive networks. Autoencoders. Transfer learning. Keras toolset. Graph Networks.

## *Programma dettagliato*

### Lezione 1

Teoria: Introduzione al corso. Contesto storico ML.

Sessione pratica: Introduzione a python, jupyter notebook, Colab. Libreria pandas.

### Lezione 2

Teoria: Definizione ML. Regression. Classification. Supervised learning. Unsupervised learning. Reinforcement learning.

Sessione pratica: Creazione e fit di un modello di regressione lineare.

Valutazione accuratezza modello su dataset test. Visualizzazione grafica del modello.

### Lezione 3

Teoria: Modello matematico. Inference e Prediction. Modelli parametrici e non-parametrici. Overfitting e underfitting. Ottimizzazione varianza e bias. Generalisation error. k-Nearest Neighbours.

Sessione pratica: Esempio di k-Nearest Neighbours. Creare predizioni con il modello ottenuto.

### Lezione 4

Training in pratica. Supervised learning. Loss function. Minimizzazione con Gradient Descent. Stochastic Gradient Descent. Modelli lineare per la classificazione. Unsupervised training (clustering) K-Means .

Sessione pratica: Esempio di KMeans con due variabili e con una variabile, confrontando le accuratezze.

### Lezione 5

Ripasso concetti propedeutici di algebra lineare. Risoluzione problema lineare.

Giustificazione geometrica dei residui quadrati.

Sessione pratica: Gradient Descent

### Lezione 6

Tree decisionali (decision trees) per regressione e classificazione. Interpretazione diagrammatica e grafica. Come costruire un tree (recursive binary splitting). Gini Index. Tree pruning. Cost complexity pruning.

Sessione pratica: Decision Tree Classifier, Pruning con scikit-learn Iris Dataset e Cancer Dataset

### Lezione 7

Decision Tree, pro e contro. Ensemble training methods. Voting classifier. Hard voting. Bootstrap. Bagging (e Pasting). Random Forest. Boosting.

Sessione pratica: Bagging, Random Forest, Boosting con scikit-learn Cancer Dataset

### Lezione 8

Artificial Neural Networks. Multilayer Perceptron.

Sessione pratica: Boosted decision tree con dati bosone di Higgs.

## Lezione 9

Convolutional Neural Networks.

Sessione pratica: Un semplice Perceptron and e una DNN sul dataset IRIS.

## Lezione 10

Confusion Matrix. ROC curve. AUC.

Sessione pratica: Convolutional Neural Network su MNIST dataset.

## Lezione 11

CNN in Fisica. Recurrent Neural Network. Auto-encoders (stacked, variational, recurrent, convolutional). Generative Adversarial Networks. Generative Models.

Sessione pratica: Auto-encoders e stacked auto-encoders.

## Lezione 12

Cenni alle Graphs Neural Networks.

Sessione pratica: Recurrent Neural Network su una serie di dati sintetica.

## **Metodologia**

Lezioni o videolezioni (a seconda della situazione sanitaria) frontali teoriche e in cui si affronta un problema mostrando come esercizio lo sviluppo di un particolare algoritmo, la realizzazione di un programma o l'analisi di un set di dati con una particolare tecnica. Inoltre saranno fatte delle sessioni di hands-on utilizzando strumenti di cloud computing in cui gli studenti dovranno sviluppare sul proprio computer alcuni gli esercizi proposti.

Modalita' di verifica delle conoscenze e capacita':

Le capacita' saranno verificate richiedendo allo studente lo sviluppo di un progetto di software per analisi dati e/o l'analisi di un set di dati attraverso le tecniche apprese nel corso. Esame orale con domande specifiche partendo da un progetto di calcolo scientifico realizzato e presentato dagli studenti

## **Materiale**

### *Testi su Machine Learning*

- Introduction to Statistical Learning. G. James. Springer edition. Ottimo libro introduttivo per la teoria (e' una sintesi del testo T).
- The Elements of Statistical Learning. T. Hastie R. Tibshirani J. Friedman - (Second edition) Springer edition. Testo di approfondimento teoria.
- Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow - Aurelien Geron - O'Reilly (2nd edition) Testo piu pratico con molti esempi (usando sci-kit) e spiegazioni intuitive.

- (Gratis online - base):  
[https://python-course.eu/books/bernd\\_klein\\_python\\_and\\_machine\\_learning\\_a4.pdf](https://python-course.eu/books/bernd_klein_python_and_machine_learning_a4.pdf)  
Testo con molti esempi dettagliati in python sia di data-visualization che di Machine Learning.
- (Gratis online - avanzato): <https://www.deeplearningbook.org/> Testo completo e dettagliato per capire profondamente il deep learning.

### *Programmazione*

- <https://docs.python.org/3/tutorial/>
- <https://www.learncpp.com/cpp-tutorial/introduction-to-cplusplus/>