

A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is a light green color. They are positioned diagonally, with the blue one in front of the green one.

# Python per il Calcolo Scientifico

Angelo Cardellicchio



# Corollario a Scikit Learn

- Encoding delle feature
- Grid search
- Pipeline



# Encoding delle feature

- Non tutte le feature sono di tipo numerico: molte, infatti, sono *categorical feature*
- Molti stimatori, tuttavia, hanno bisogno di feature di questo tipo per funzionare in maniera adeguata!
- Esistono diversi *encoder* che ci permettono di trasformare le categorical feature

Feature	Ordinal Encoder	One-Hot Encoder
Male	1	1, 0
Female	2	0, 1

# Grid search

- La *grid search* prevede una ricerca 'a griglia' dei valori ottimali degli iperparametri
- Ad esempio, per un albero di regressione, potremmo voler valutare come variano le prestazioni al variare di criterio di split e massima profondità dell'albero
- Supponendo di avere MAE ed MSE come criterio di split, e 2 e 4 come possibili profondità dell'albero, la grid search sceglierà i parametri che minimizzano una metrica esplorando tutte le possibili combinazioni degli iperparametri

**2, MSE**

**4, MSE**

**2, MAE**

**4, MAE**



**Evaluate**

**2, MAE**

# Pipeline

- Come abbiamo visto, la maggior parte degli algoritmi di machine learning non può essere usato in isolamento, ma piuttosto come parte di una *sequenza algoritmica*
- Questa sequenza è comunemente nota come **pipeline**
- Ad esempio, una pipeline può essere composta da un algoritmo di normalizzazione, cui segue una PCA, cui segue il Kmeans
- Scikit-Learn ci offre una classe di oggetti che ci permette di ottenere in modo rapido e semplice questo risultato

