Advanced School in Artificial Intelligence

Introduction to Machine Learning

Elena Bellodi

elena.bellodi@unife.it

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Regione Emilia-Romagna

FERRARIAE UNIVERSITAS · 13 91 · EX LABORE FRUCTUS

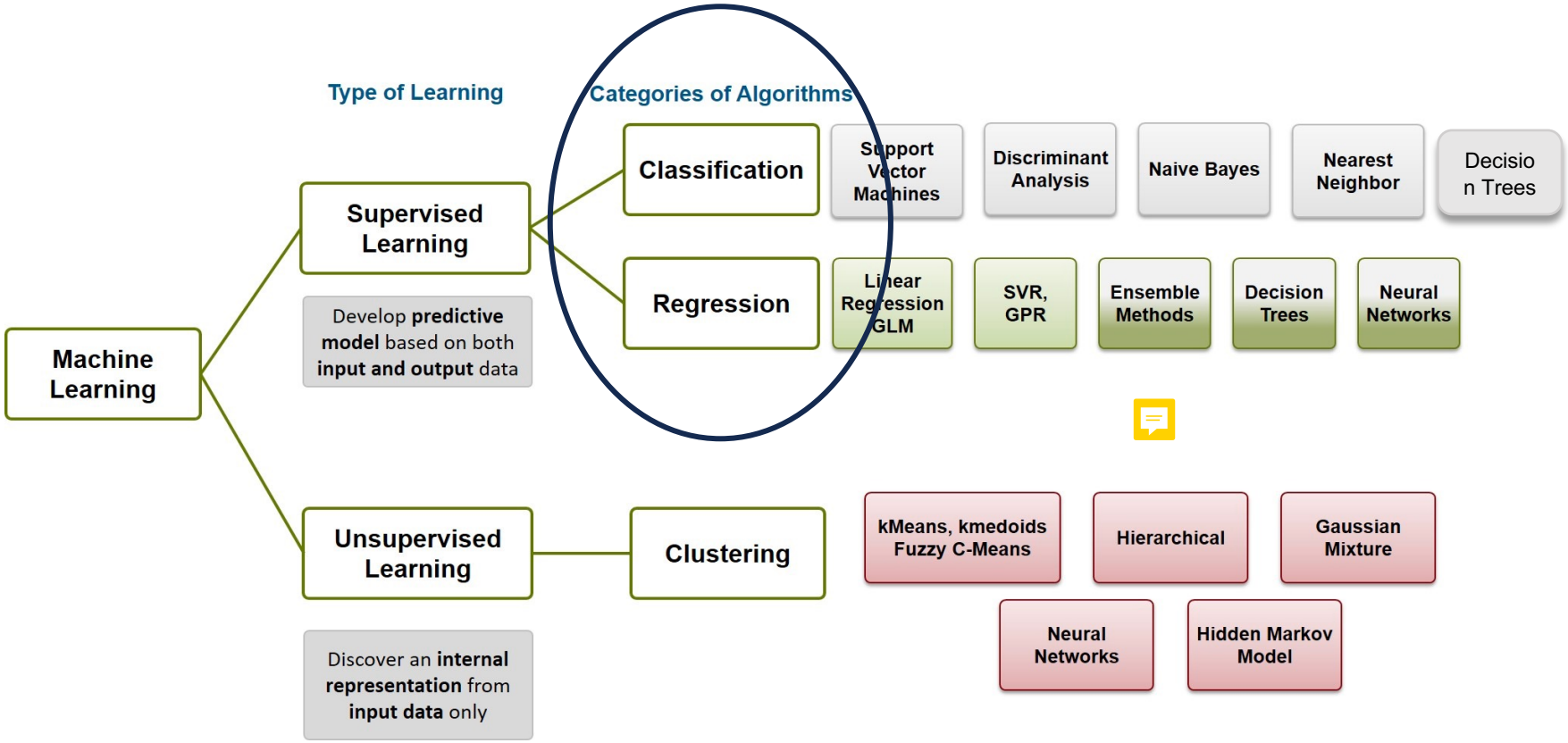
Università degli Studi di Ferrara

Outline

- Machine learning (ML) definitions
- Learning paradigms
  - **supervised**
  - **unsupervised**
  - semi-supervised
  - reinforcement
- Use of Data in ML
  - training, validation and test set
  - generalization, underfitting and overfitting
  - capacity
  - bias and variance
- Learning protocols

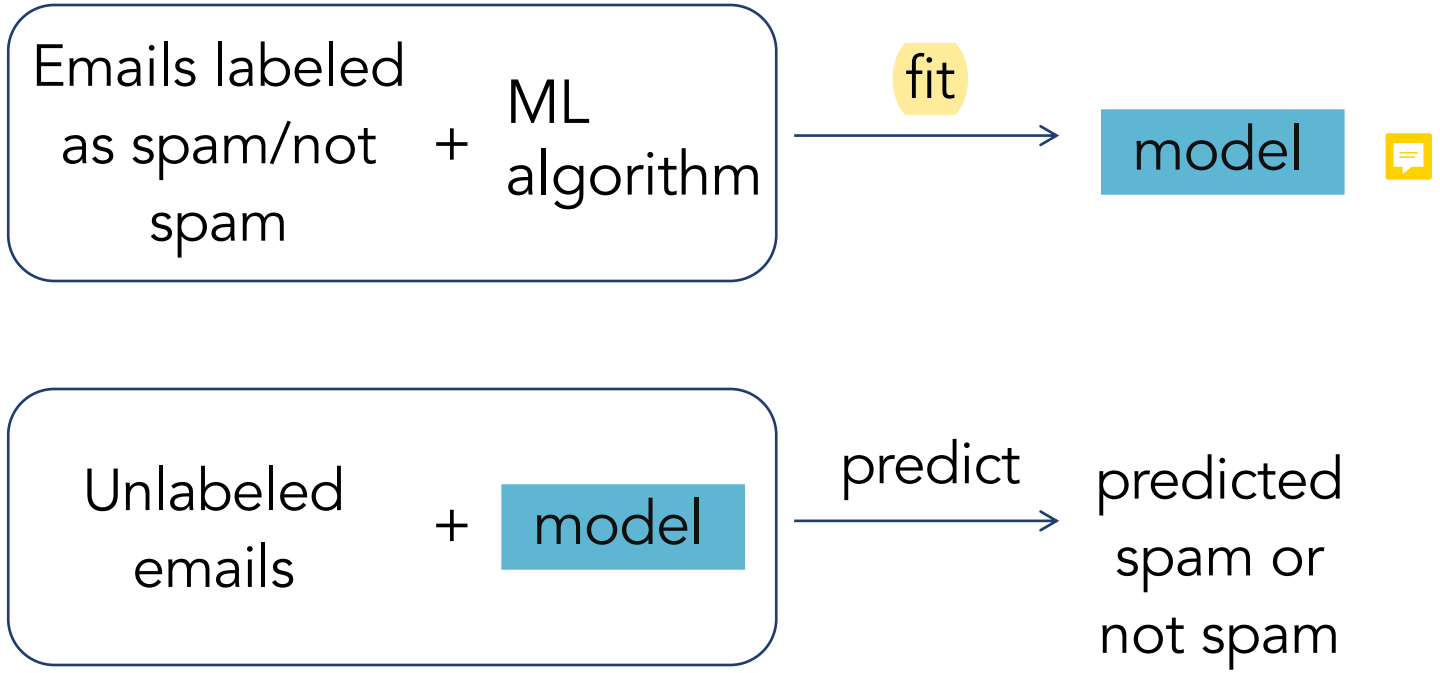
Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Supervised Learning Algorithms



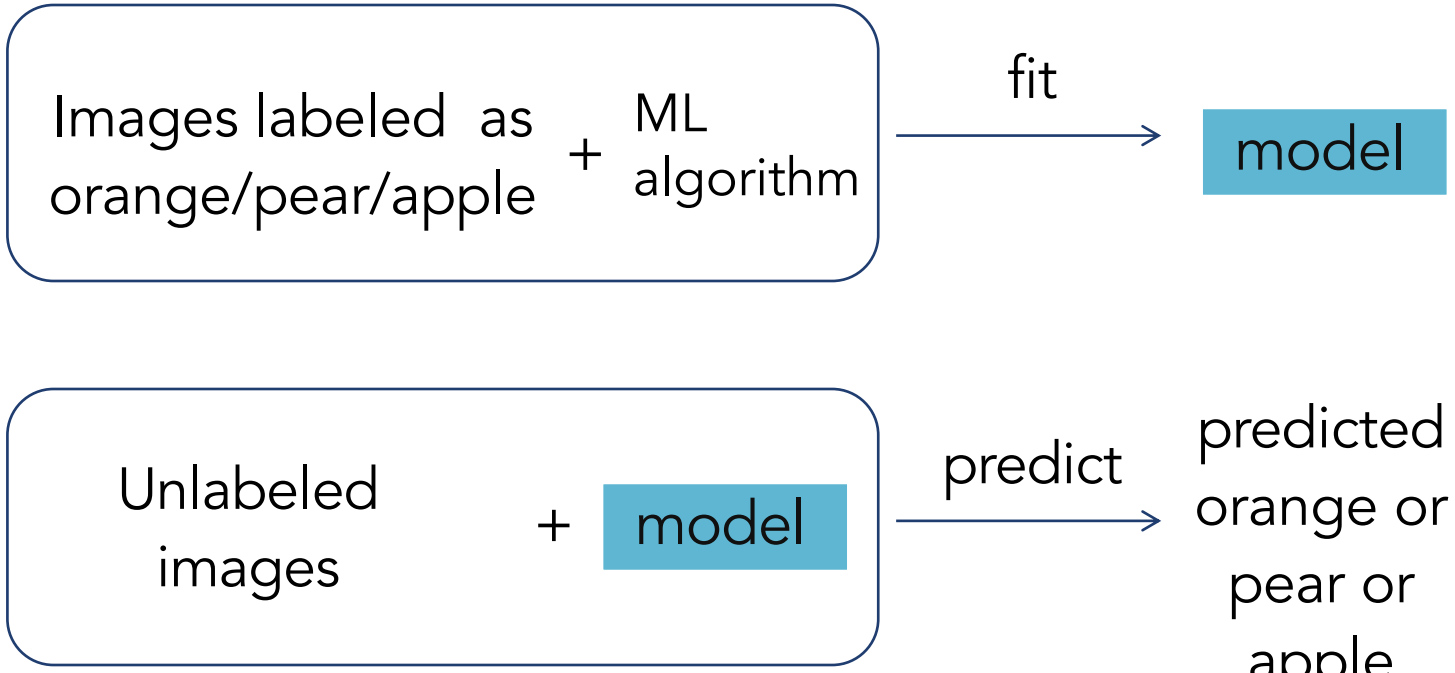
Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Classification: binary case



Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Classification: multiclass case



Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

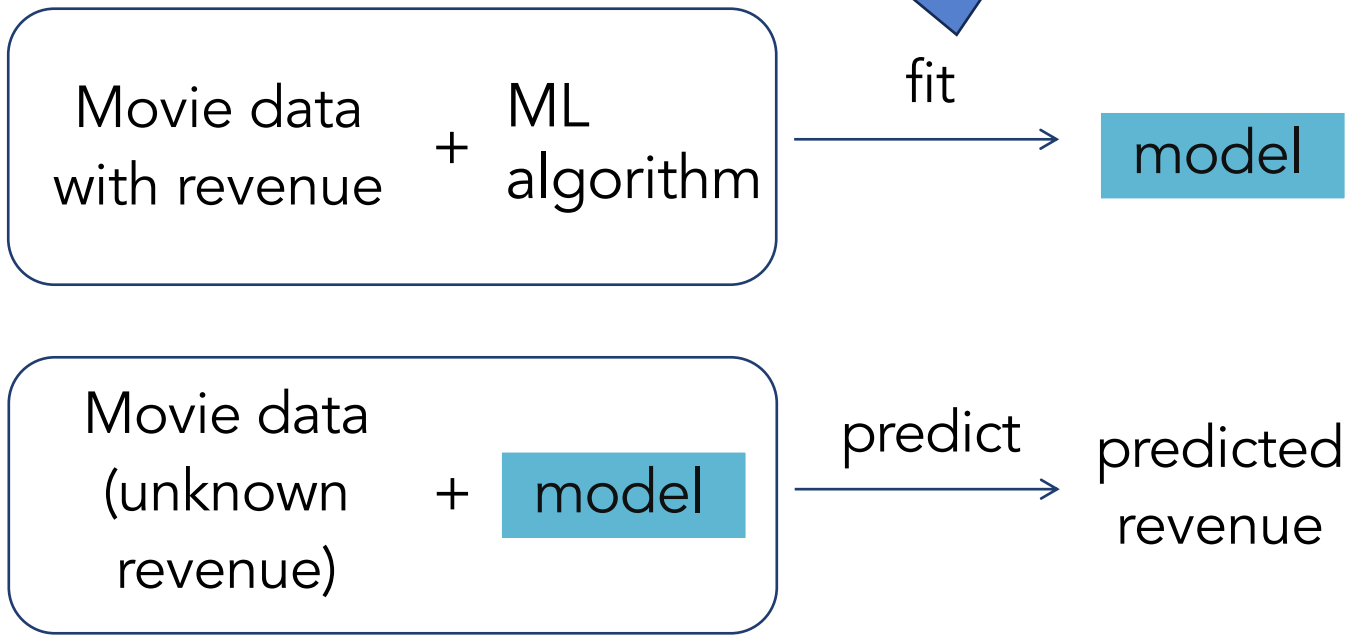
Classification

- We are going to study Tree Models for classification tasks

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Regression

The ML algorithm, surveying the past movie data **x**, will fit an 'estimator function' to this data to learn y (the revenue) as a function of the movie attributes



Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Regression

- Linear regression: a linear model, meaning that g is a linear combination of input features, with weights applied to each feature
  - for a single feature

$$\hat{y}_i = g(x) = w_0 + w_1 x$$

- Polinomial regression

$$\hat{y}_i = \mathbf{w}^{\mathbf{T}}\mathbf{x} = \mathbf{w} \cdot \mathbf{x} = w_0 + w_1 x + w_2 x^2 \ldots + w_k x^k$$

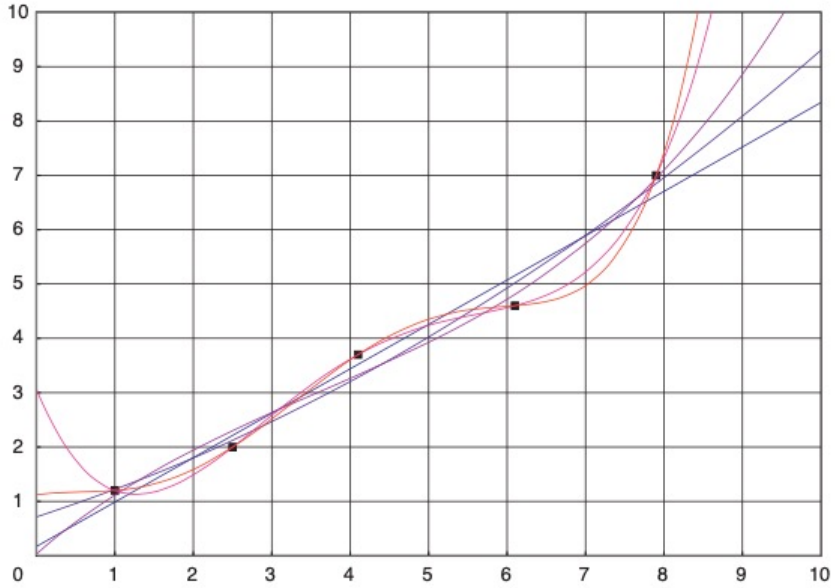
- Logistic regression

$$\hat{y}_i = \sigma(\mathbf{w}^{\mathbf{T}}\mathbf{x})$$

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Regression

- The output and the inputs - attribute values - are all numeric
- The problem is to come up with good values for the **weights** - ones that make the model's output match the desired output
- **Linear and polynomial models** are the easiest to **visualize in two dimensions**, where they correspond to drawing a curve through a set of data points, which represents the prediction equation

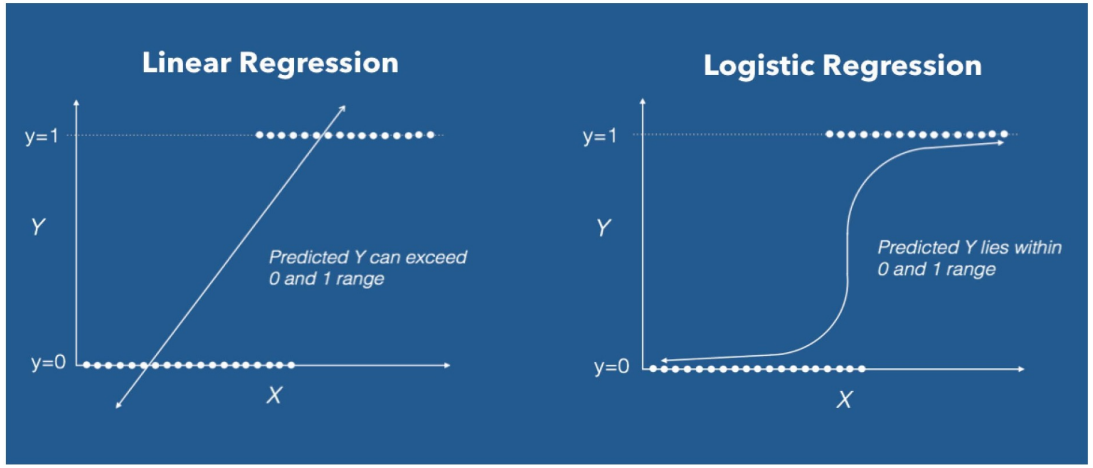


Polynomials of different degree fitted to a set of 5 points. From bottom to top in the top right-hand corner: degree 1 (straight line), degree 2 (parabola), degree 3, degree 4, degree 5

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Regression

- **Logistic** regression is a linear model that outputs a probability, a value **between 0 and 1**
- The output can be interpreted as a probability for a binary event (heart attack or no heart attack, etc.), so the final prediction is discrete
- $\sigma(s) = \frac{e^s}{1-e^s}$ is the logistic function whose output is between 0 and 1



Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Fitting Training and Test Data (Supervised learning)

Training Data

Fit the model

Test Data

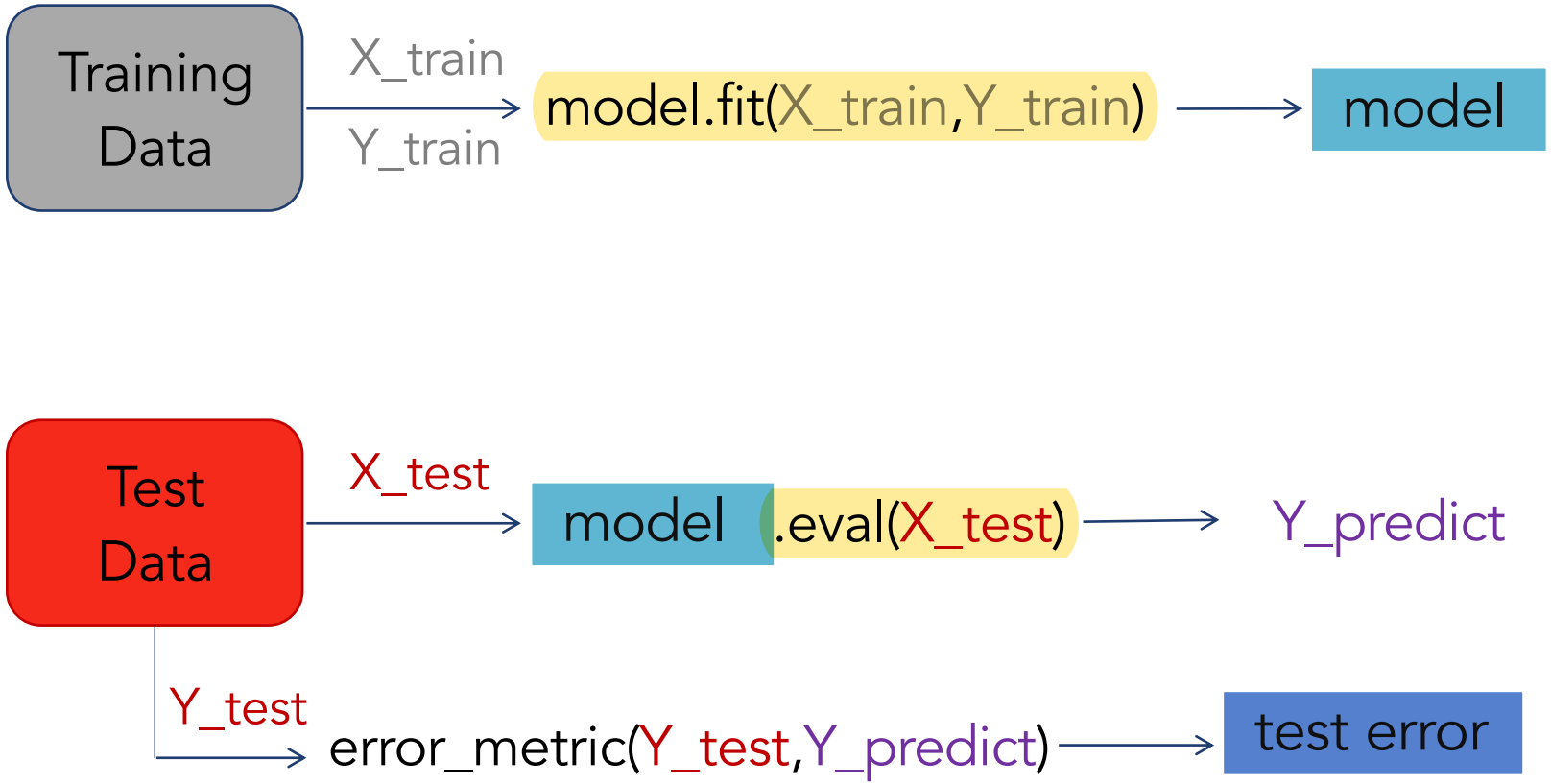
Measure performance

- Predict label with model
- Compare with actual value
- Measure error

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

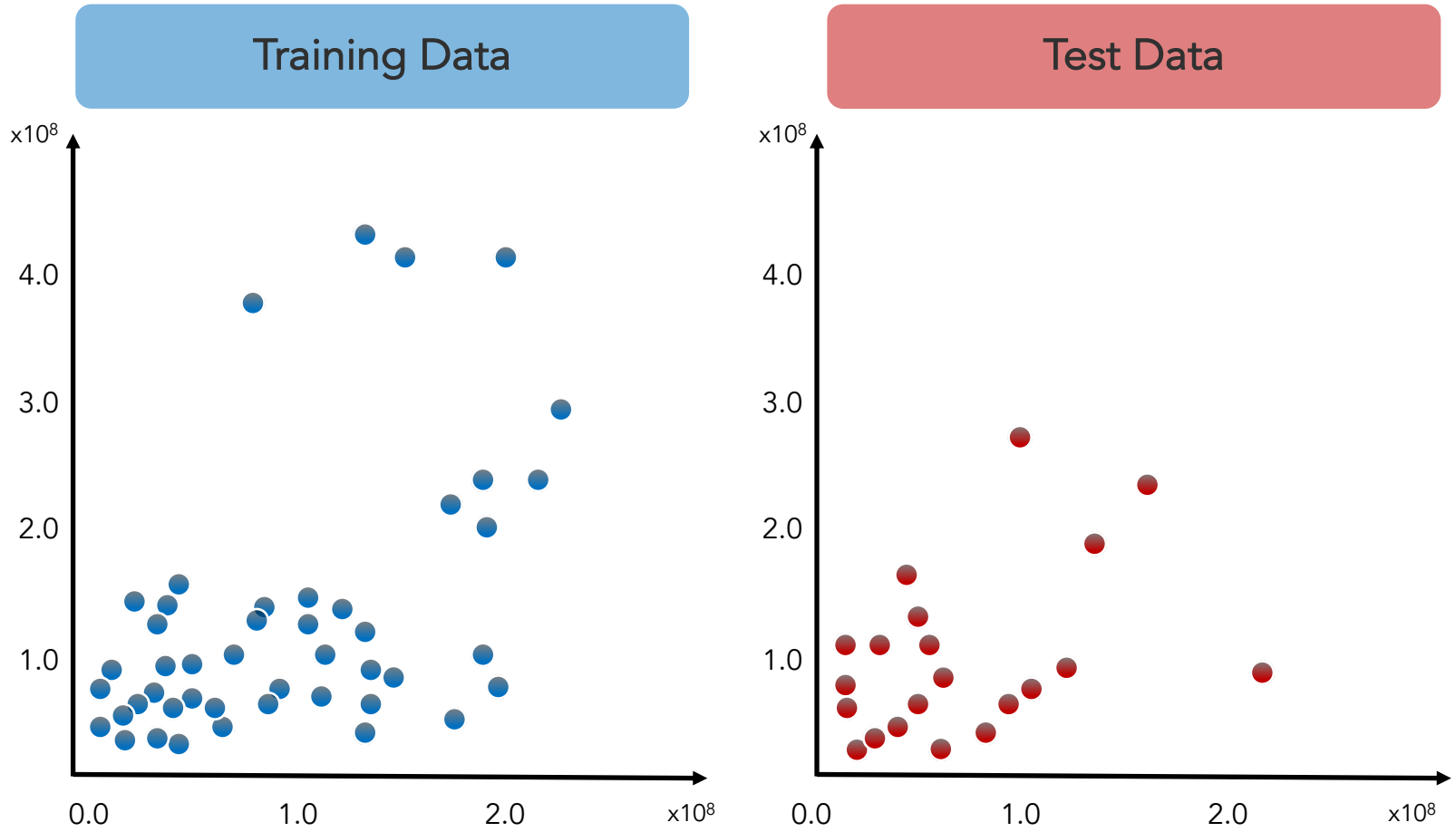
Fitting Training and Test Data

(no cross-validation)



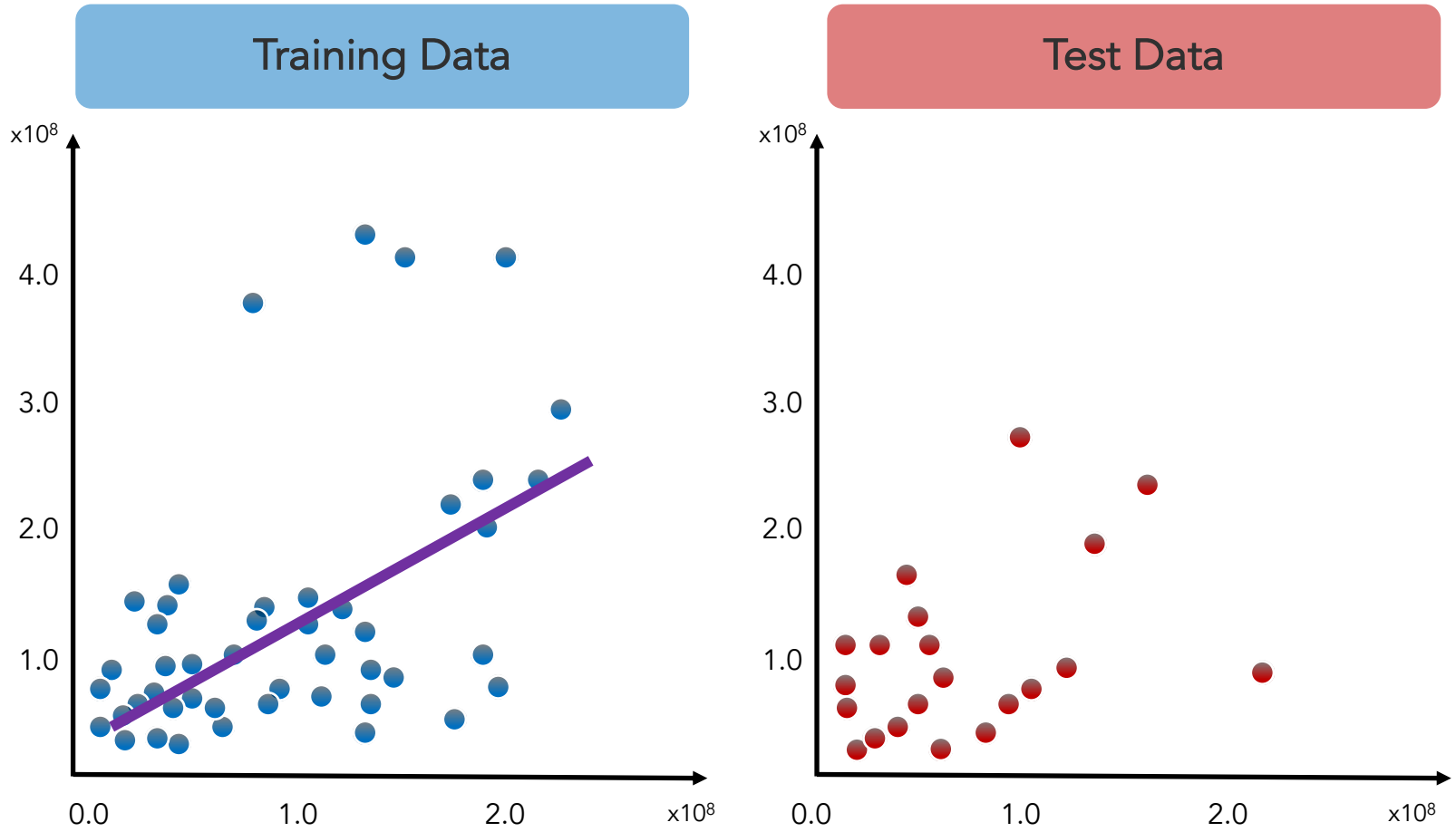
Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Fitting Training and Test Data



Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

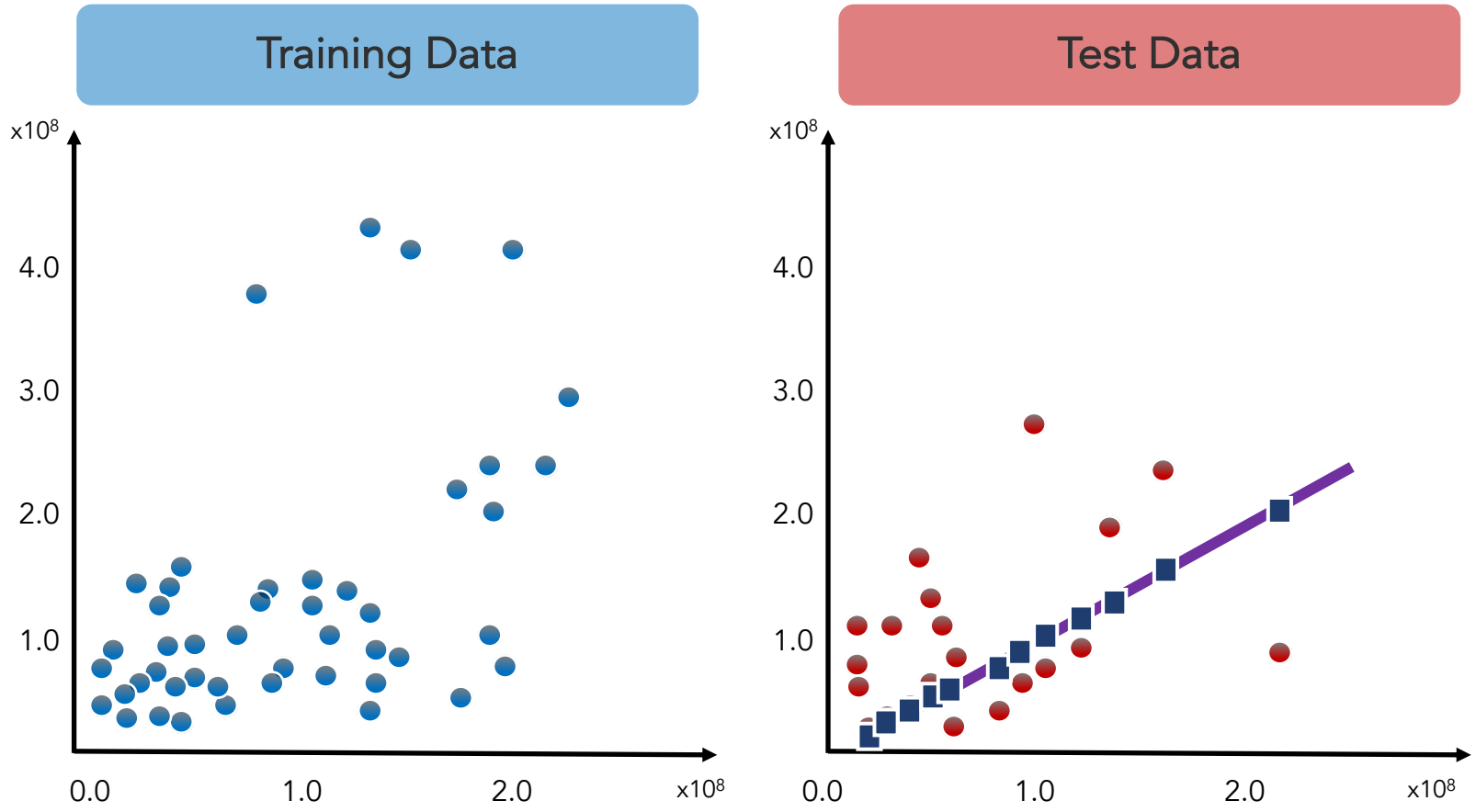
Fitting Training and Test Data



Fit the model

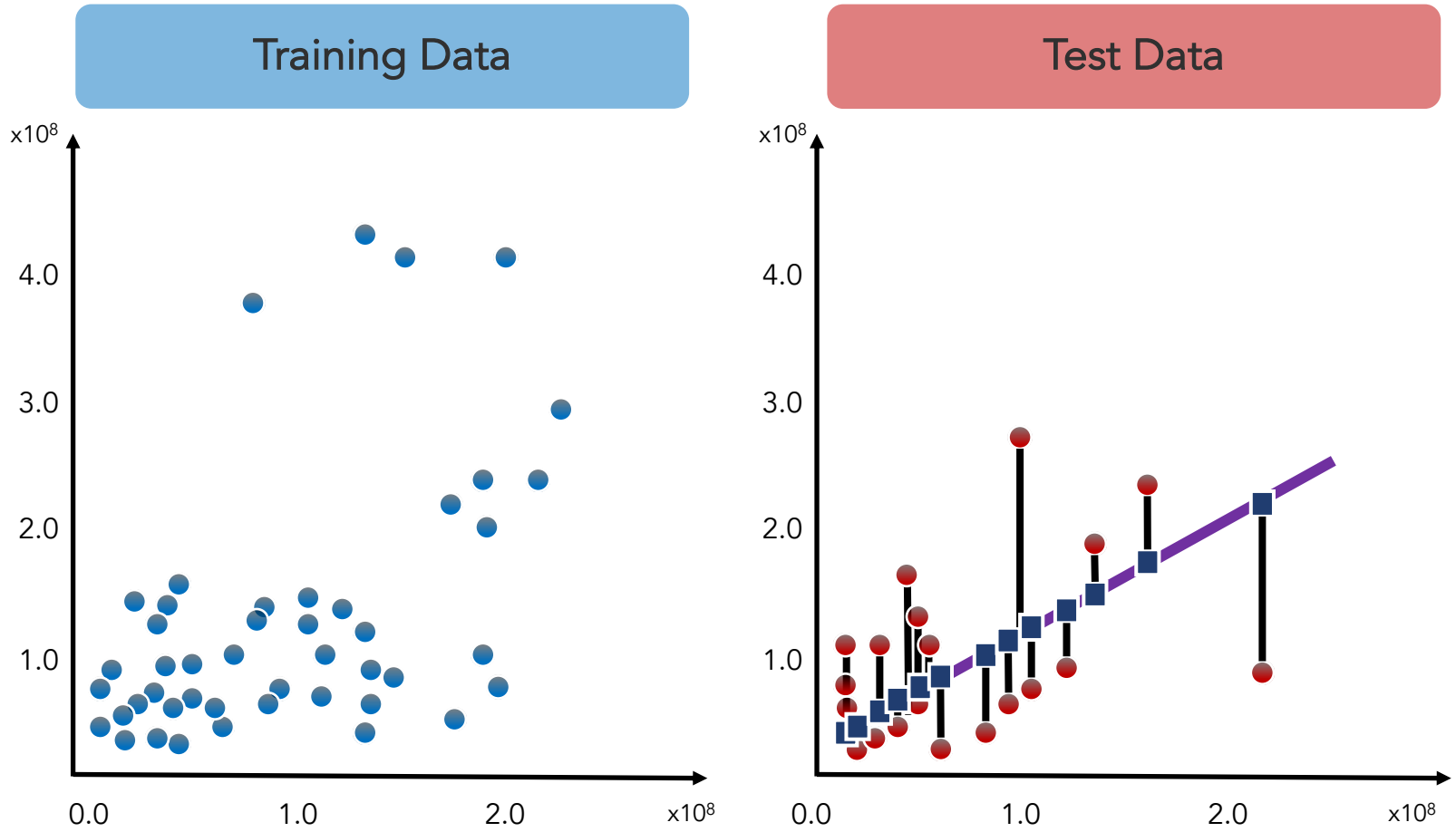
Progetto di alta formazione in ambito tecnologico, economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Fitting Training and Test Data



Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Fitting Training and Test Data



Measure error

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021