Hello!

Name: Dov Danon
Email: dov84d@gmail.com
PH.D student under supervision of Prof. Cohen-Or

Ask questions

Introduction to machine learning

Arthur Samuel (1959). Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.

What is Computer Science?

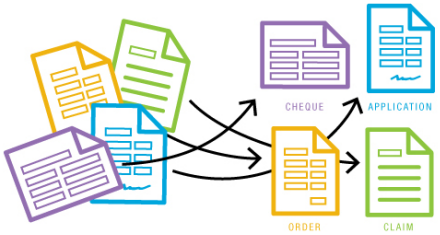
- Input-Output

What is Machine Learning?

- Input- training examples
  Output – Algorithms that can act "correctly" on new unseen examples

Tasks in machine learning

Document or photo classification



Speech recognition

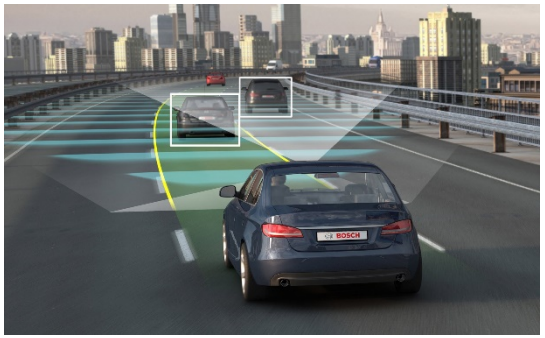
Language Translation

Tasks in machine learning

Recommendation systems

Fraud detection

Driving a car



Tasks in machine learning

Spam Filter

Basic assumption

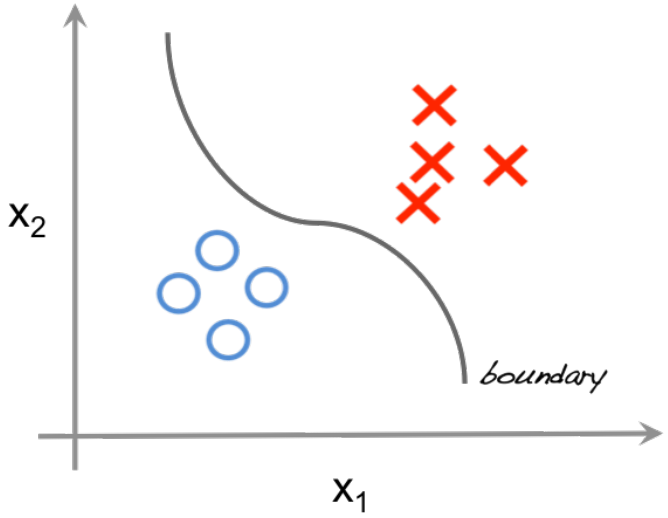
Is the past really represent the future?

YES!

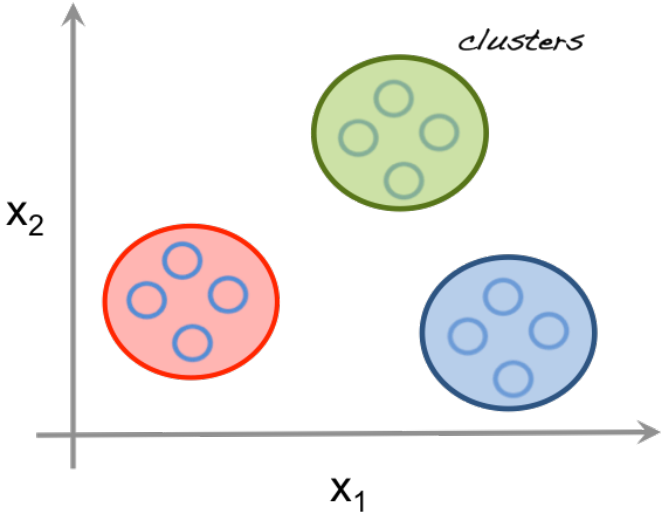
IID:
Independent and Identically Distributed

Supervised vs Unsupervised

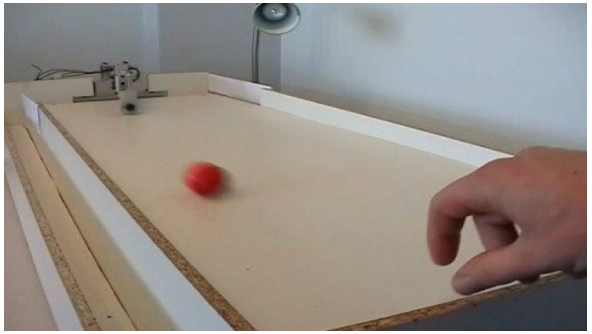
Supervised learning

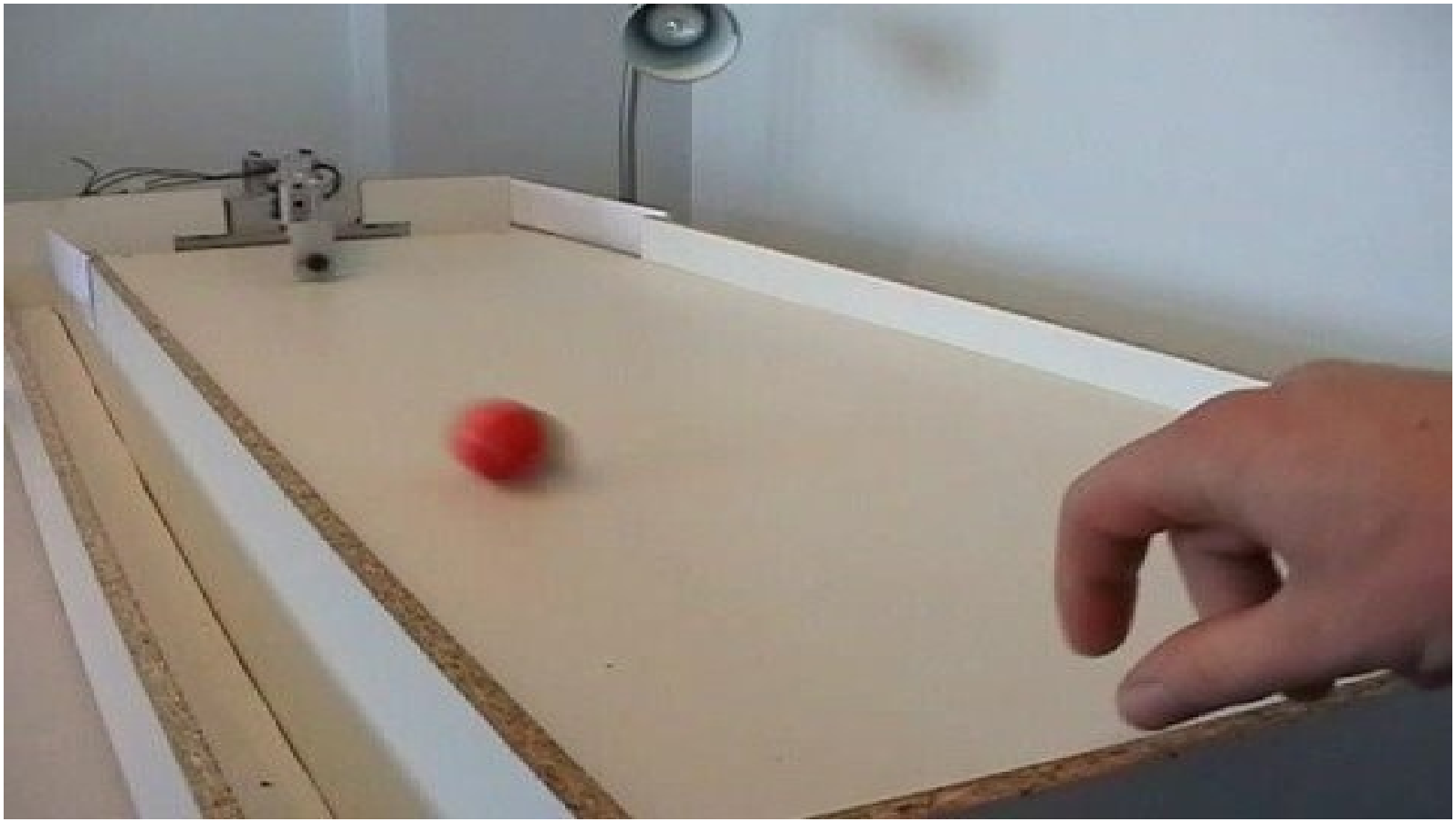


Unsupervised learning

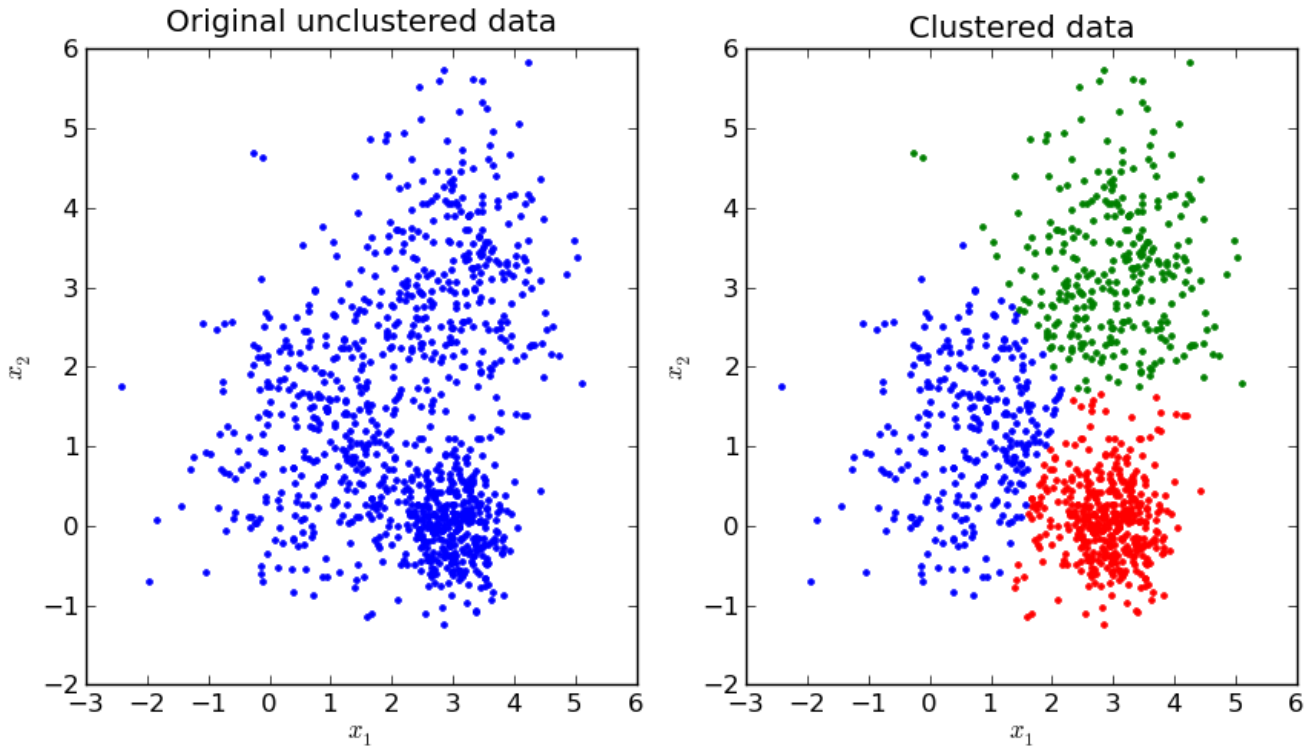


Others: Reinforcement learning





K-Means - Unsupervised



K-Means - unsupervised

Minimize:

$$\sum_{i=1}^{k} \sum_{x_j \in S_i} \|x_j - \mu_i\|^2$$

NP-Hard

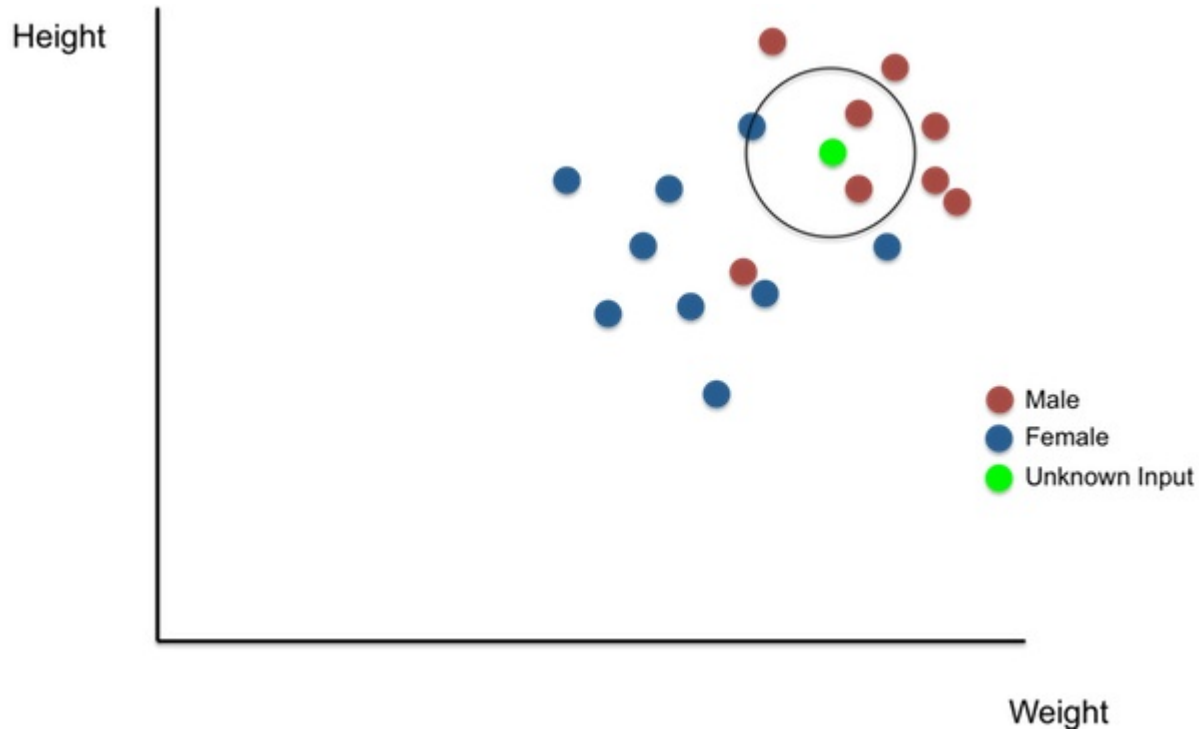
K-Means

Initialize Set $t = 1$ and select k values $\mu_1^t, \ldots, \mu_k^t$.

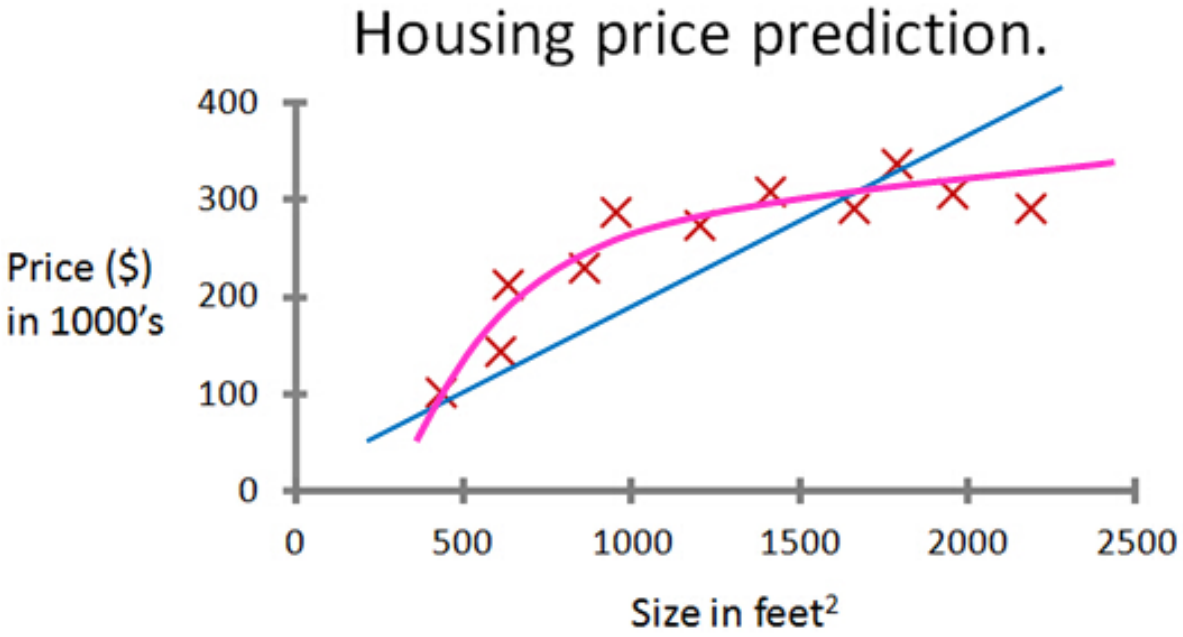
Assign Assign each x_j to the closest out of $\mu_1^t, \ldots, \mu_k^t$. Formally, $C_j^t = \arg\min_i \|x_j - \mu_i^t\|^2$ and $S_i^t = \{x_j | C_j^t = i\}$.

Update Given the sets $S_1^t, \ldots, S_k^t$ re-compute $\mu_1, \ldots, \mu_k$, by setting μ_i^{t+1} to the average in S_i^t, i.e., $\mu_i^{t+1} = (1/|S_i^t|) \sum_{x_j \in S_i^t} x_j$.

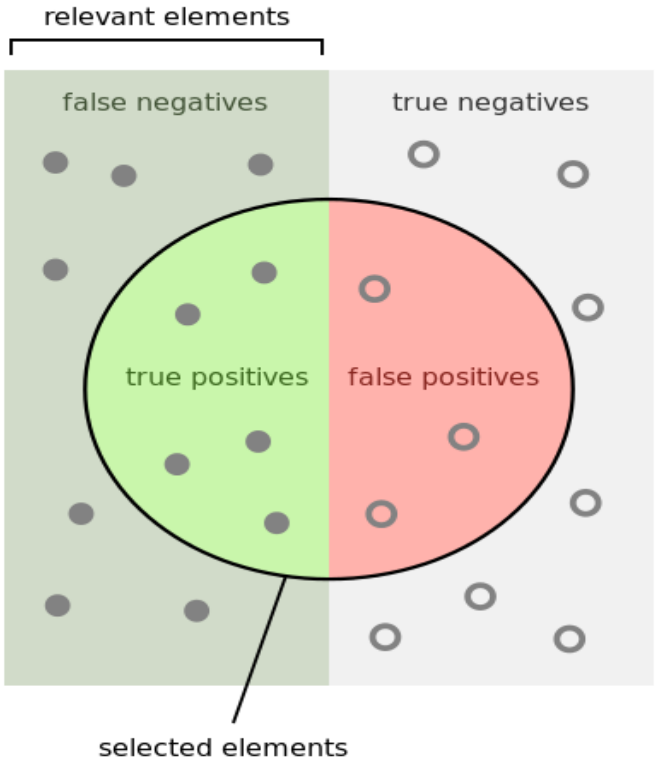
Nearest neighbor - supervised



Regression – also supervised



Classification measurements – Precision Recall



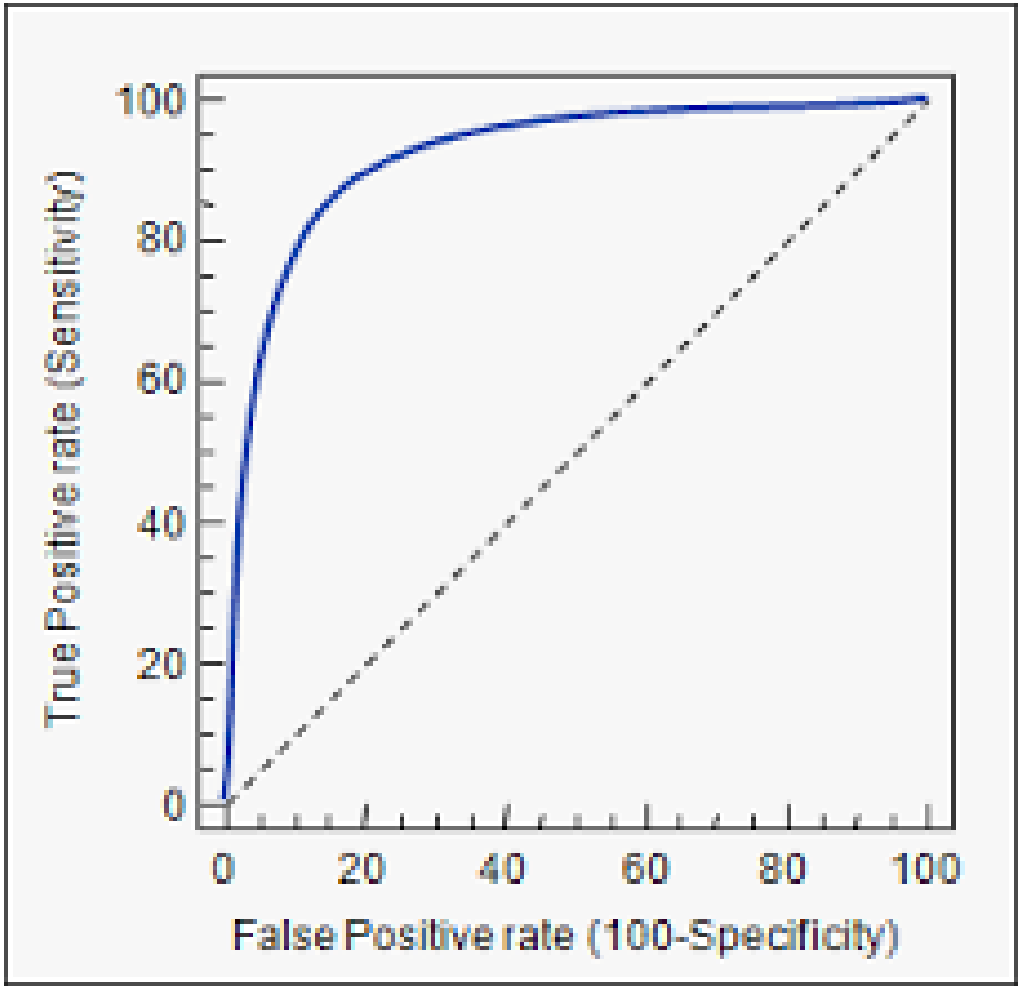
How many selected items are relevant?

$$\text{Precision} = \frac{\text{true positives}}{\text{selected elements}}$$

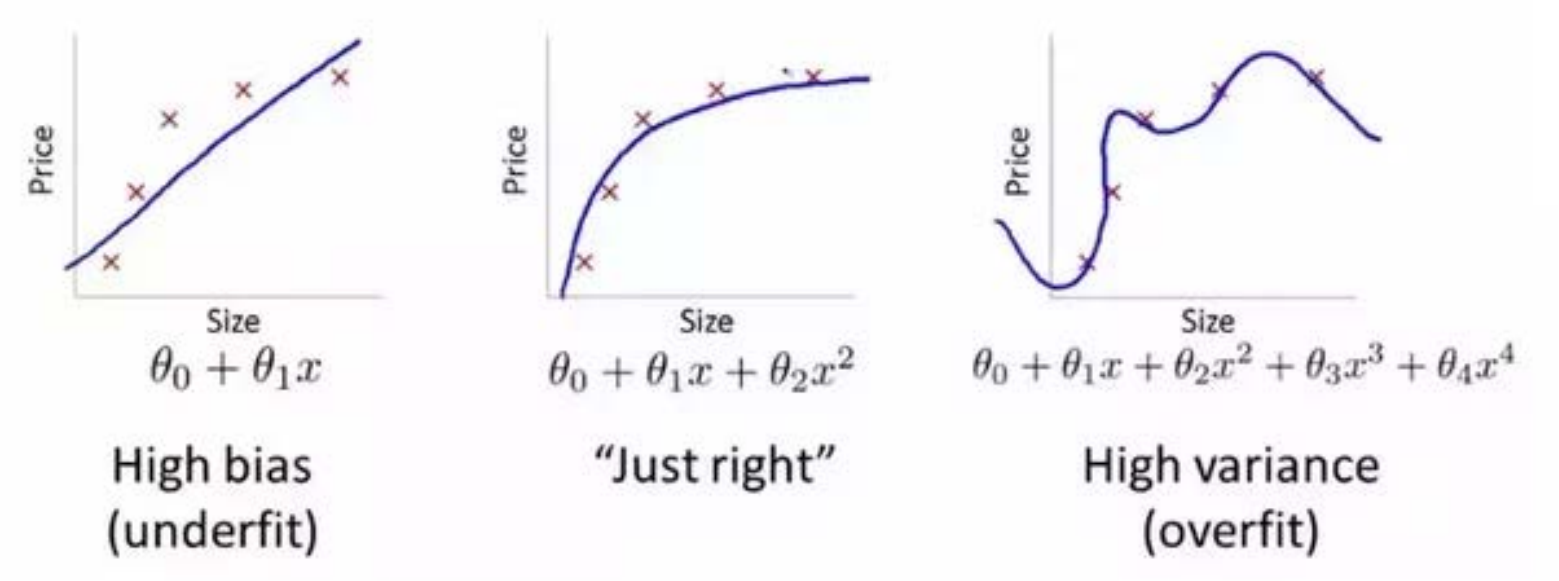
How many relevant items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{relevant elements}}$$

Classification measurements – ROC



Overfit and Uderfit

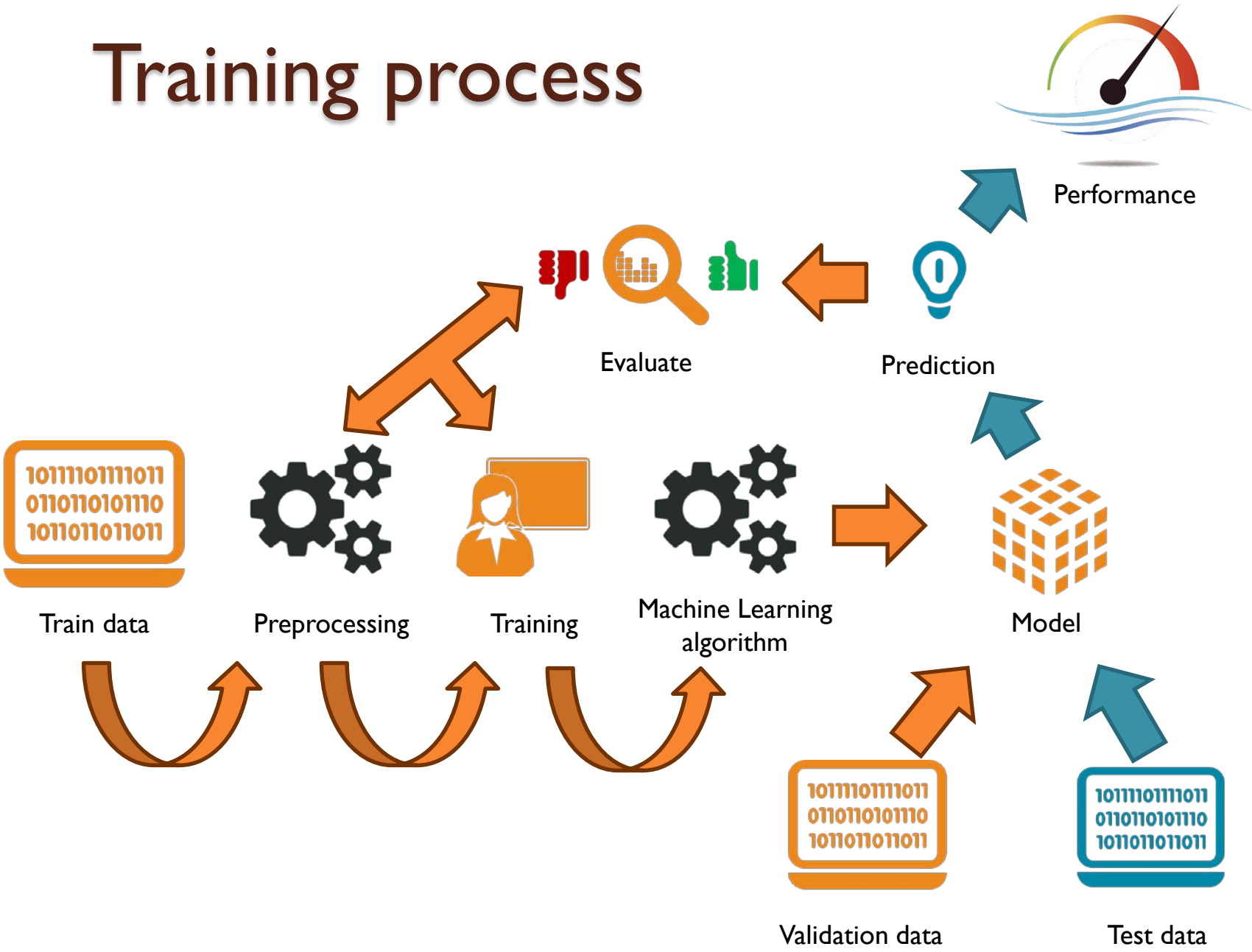


Training process

Split the data to:

- Train data – use for train the model
- Validation data – use for validate the model
- Test data – use to get the "real world" accuracy of the mode. Use only ones.

Training process




Questions