Erasmus
School of
Economics

Machine Learning

FEM31002

Introduction

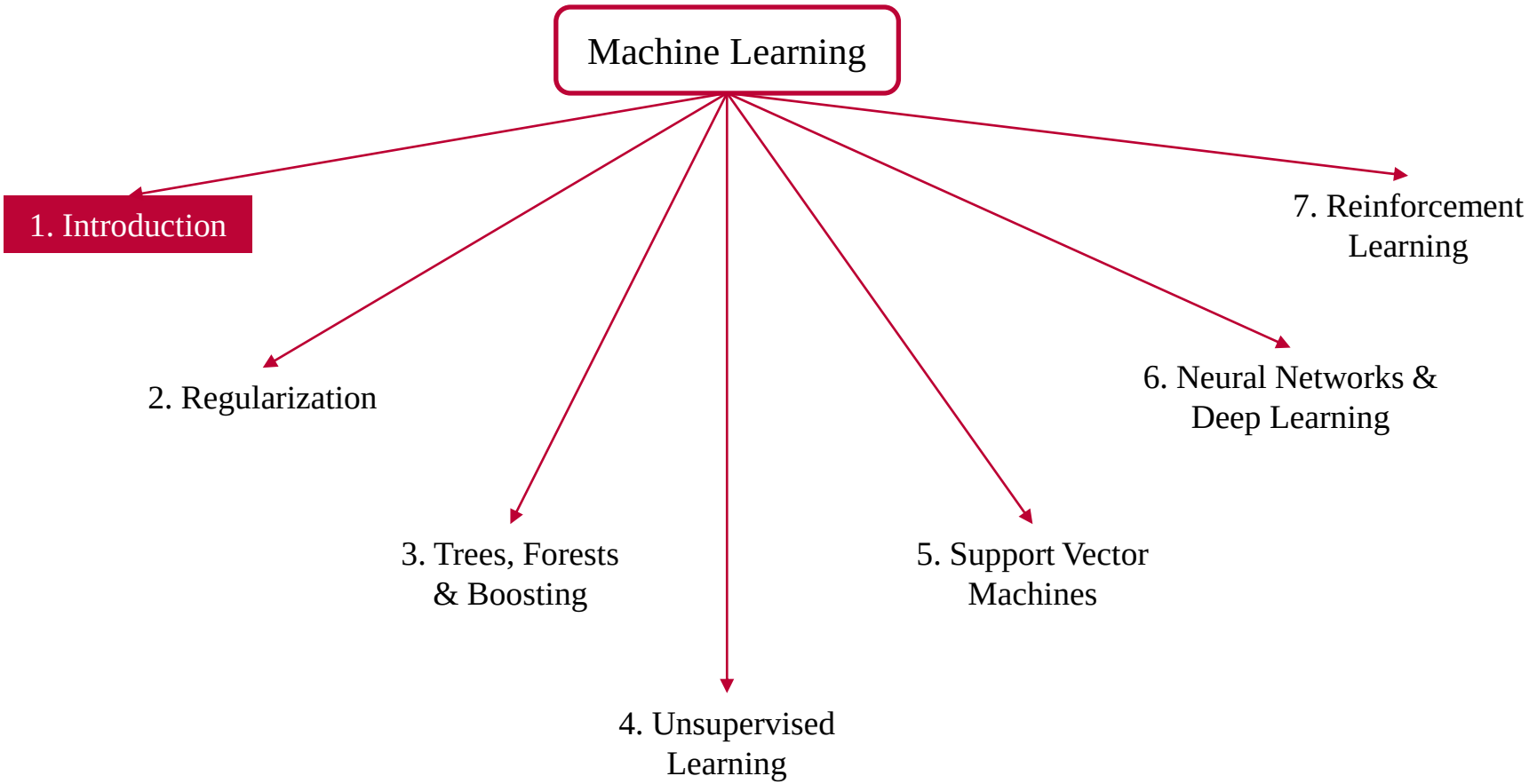
Part 1

Ilker Birbil

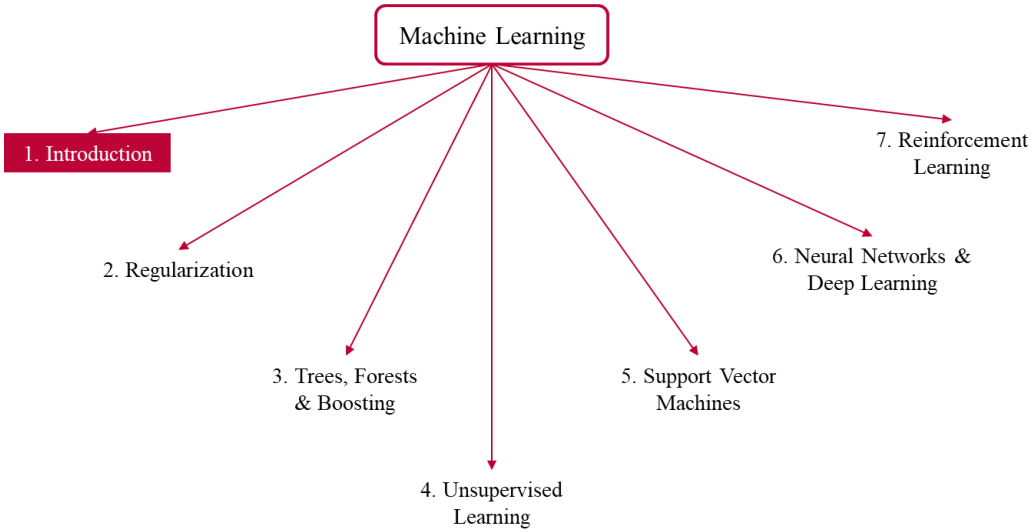
birbil@ese.eur.nl

ERASMUS UNIVERSITEIT ROTTERDAM

Outline


Machine Learning
1. Introduction
2. Regularization
3. Trees, Forests & Boosting
4. Unsupervised Learning
5. Support Vector Machines
6. Neural Networks & Deep Learning
7. Reinforcement Learning

Outline

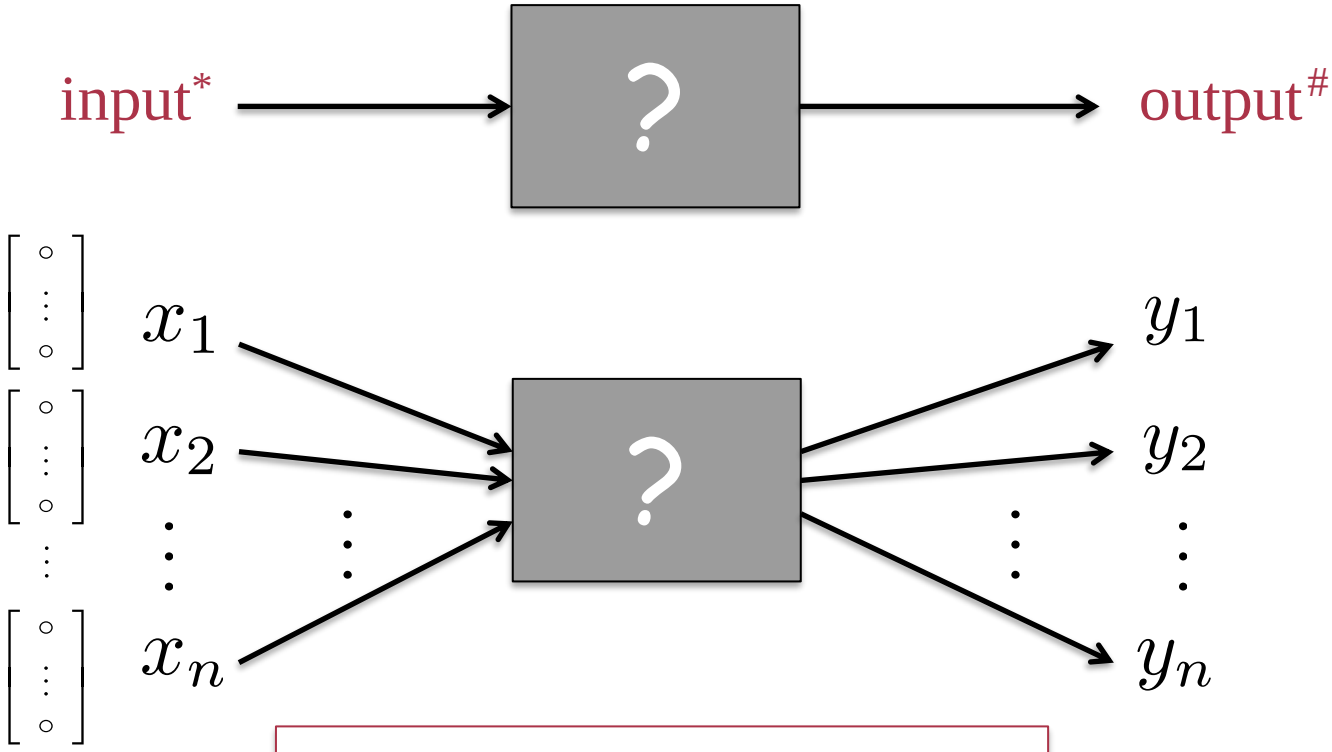


- Supervised Learning vs. Unsupervised Learning
- Train-test errors and overfitting
- Bias vs. Variance
- Bayes Classifier vs. K-Nearest Neighbor (KNN)
- Cross validation and bootstrap
- Model evaluation and algorithm comparison

Machine Learning

Statistics – Optimization – Computer Science

Supervised Learning



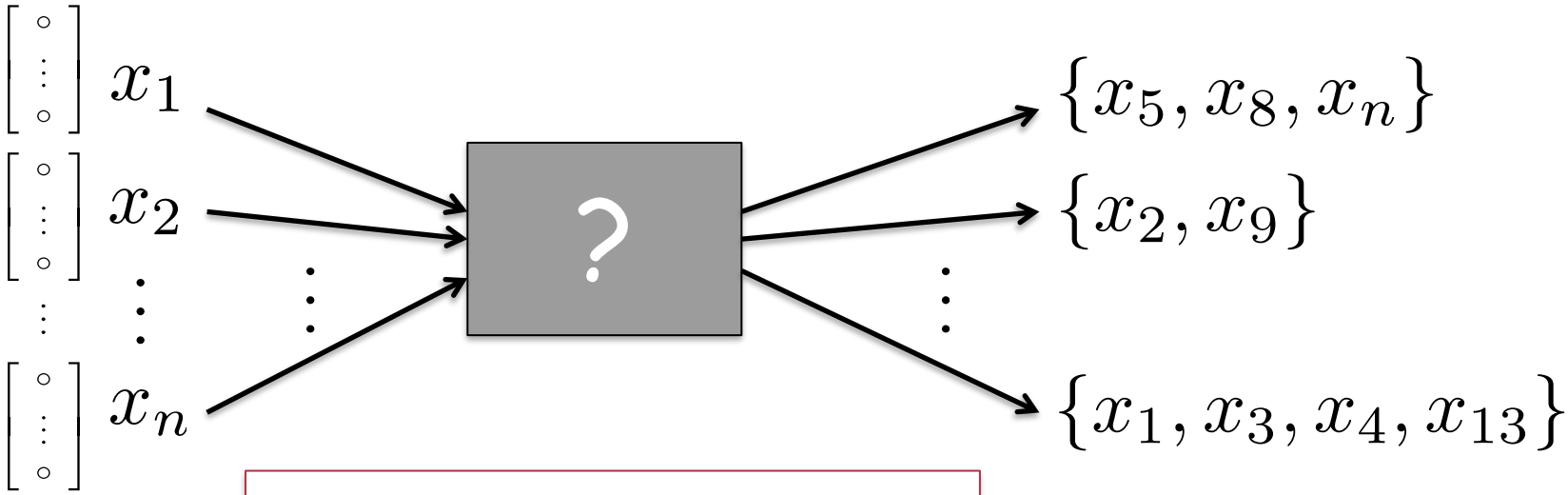
training data

$$\{(x_i, y_i) : 1, \ldots, n\}$$

* independent variable; predictor

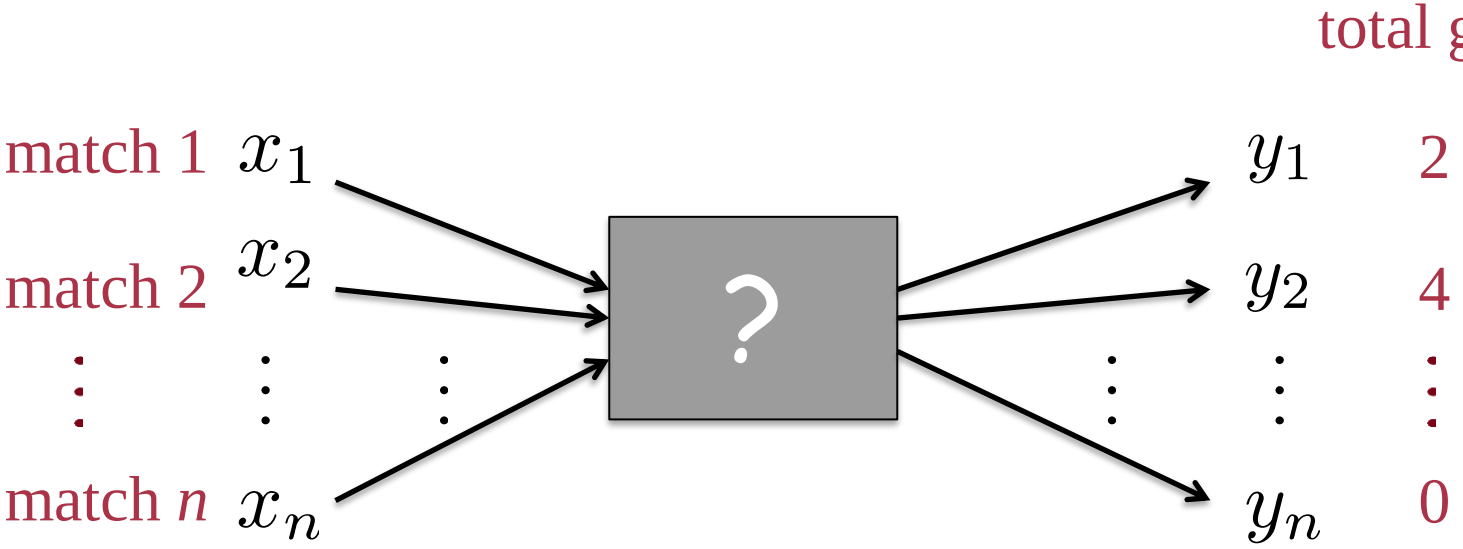
dependent variable; target value

Unsupervised Learning



training data

$$\{x_i : 1, \ldots, n\}$$



$$x_i = \begin{bmatrix} 34 \\ 27 \\ 1 \\ 0 \end{bmatrix}$$

features

total goals scored in the season

number of red cards in the season

weather: (0) rainy, (1) sunny

moral: (-2) bad, (0) normal, (2) excellent

Supervised learning

$$\{(x_i, y_i) : i = 1, \ldots, n\}$$

Spam filtering (classification)
Medical screening (classification/regression)
Shopping amount prediction (regression)

... approximation

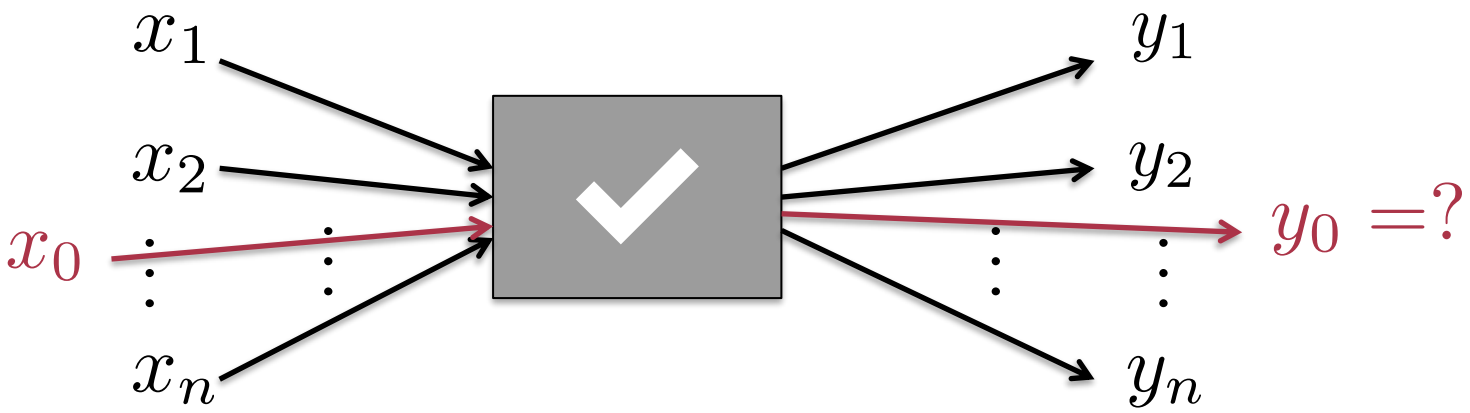
Unsupervised learning

$$\{x_i : i = 1, \ldots, n\}$$

Customer clustering (segmentation)
Fraud detection (clustering)
Dimension reduction (compression)

... description

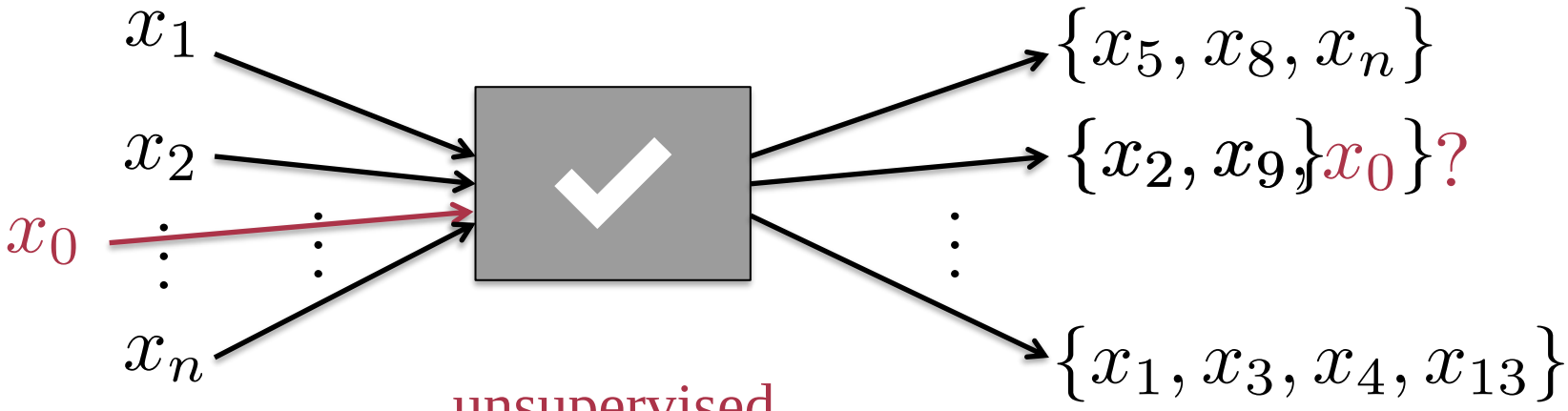
(x_0, y_0) : test data



supervised

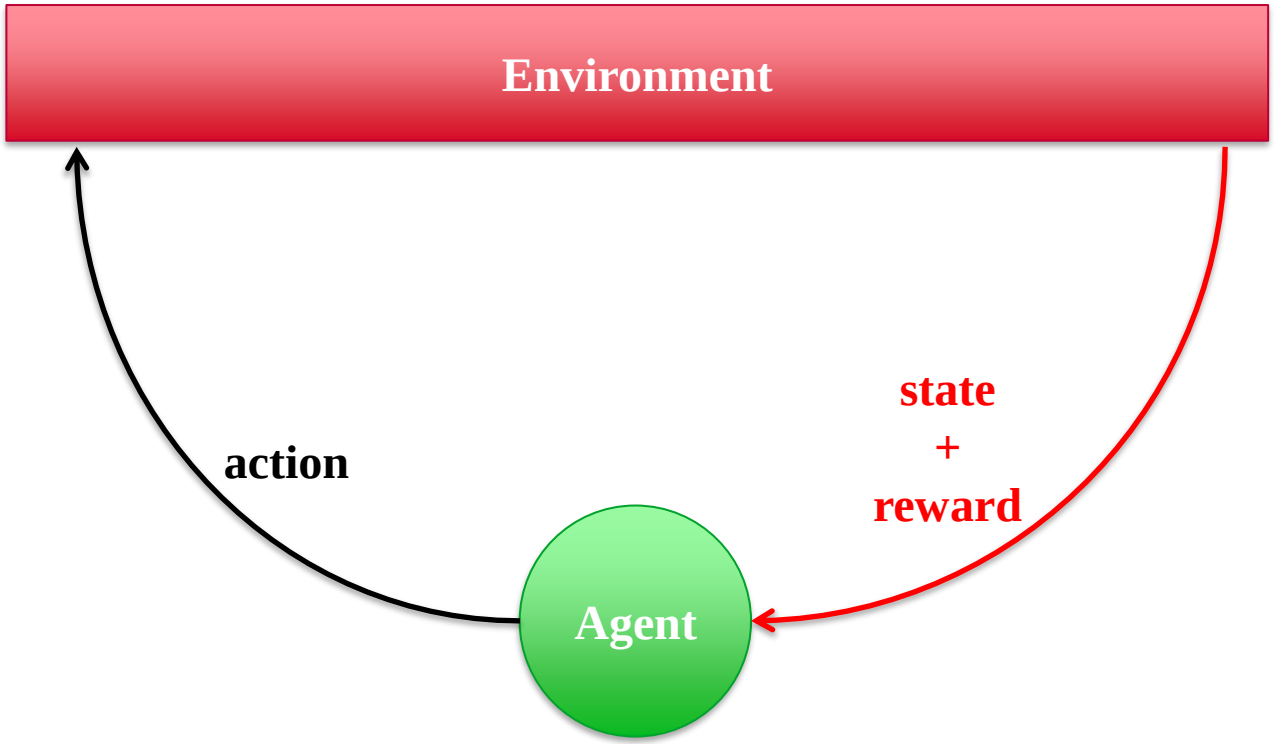
---

x_0 : test data



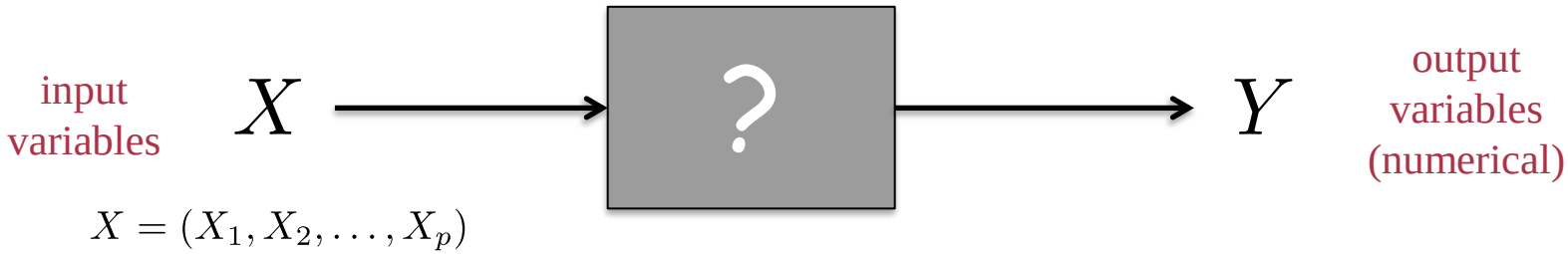
unsupervised

Reinforcement Learning



Watch the full documentary on
Netflix - AlphaGo (2017)

REGRESSION



unknown function

$$Y = f(X) + \epsilon$$

random error term

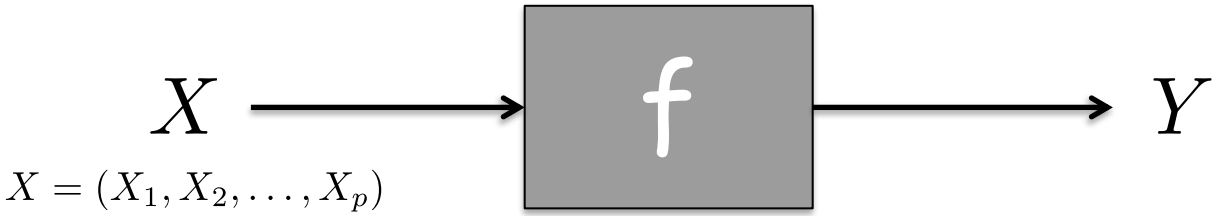
(independent of input with mean 0)

$$Y = f(X) + \epsilon \xrightarrow{\text{approximation?}} \hat{Y} = \hat{f}(X)$$

$$Y - \hat{Y} = ?$$

$\hat{f}, X$ fixed

$$\mathbb{E}(Y - \hat{Y})^2 = \underbrace{\big(f(X) - \hat{f}(X)\big)^2}_{\text{can do something (method/model)}} + \underbrace{\text{Var}(\epsilon)}_{\text{no chance}}$$

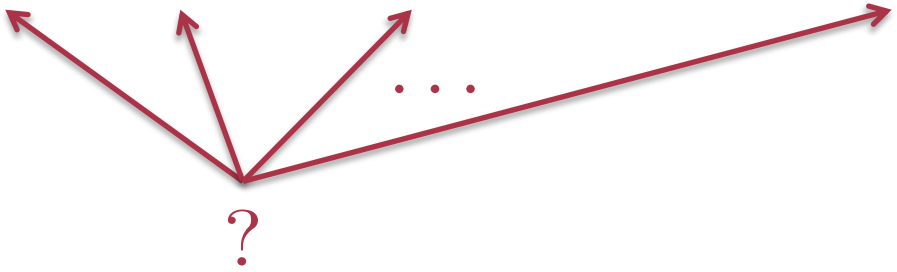


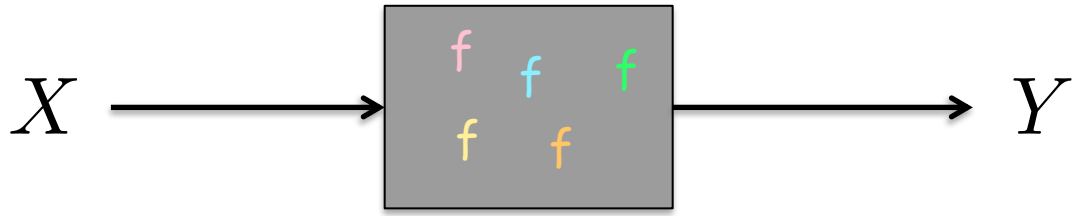
Parametric Methods

Linear Model

(assumption)

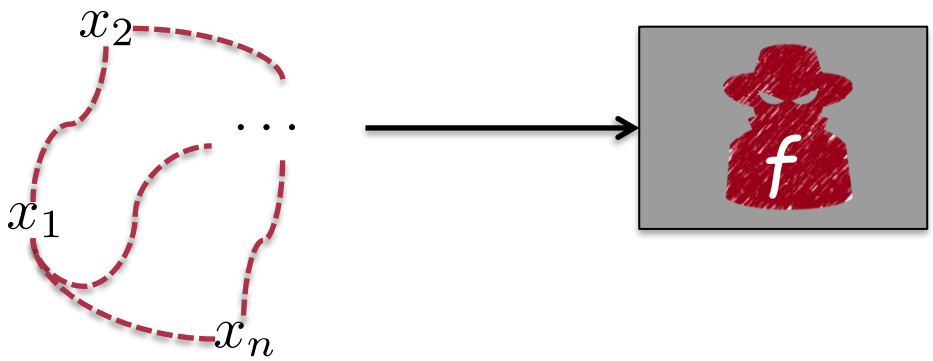
$$f(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$



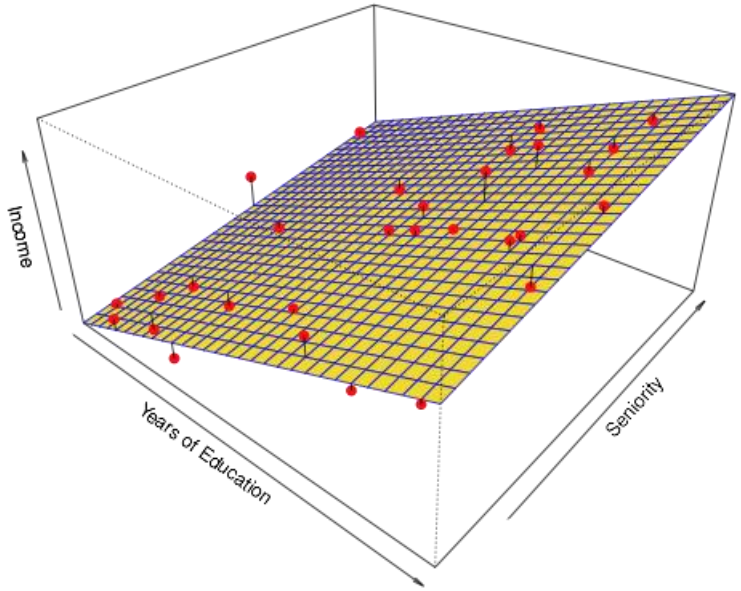


Non-parametric Methods

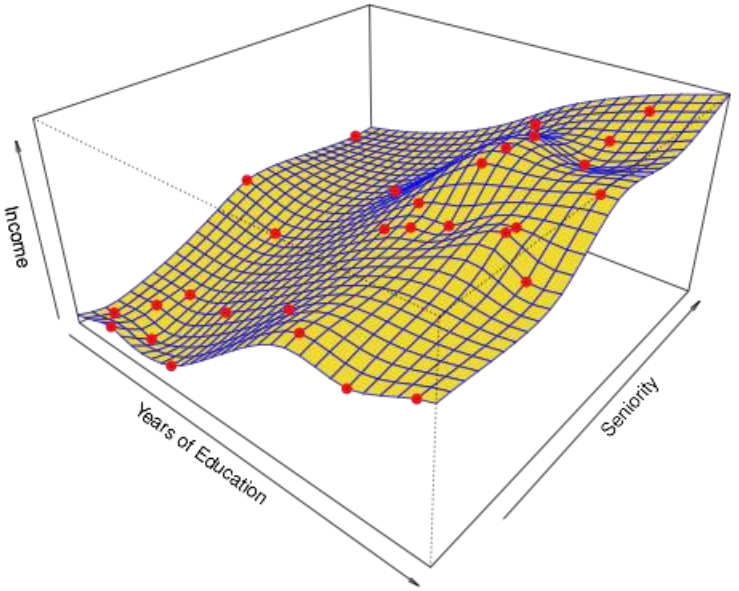
$$x_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix}$$



Parametric Method (linear model)



Non-parametric Method (spline)



$$f(X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$