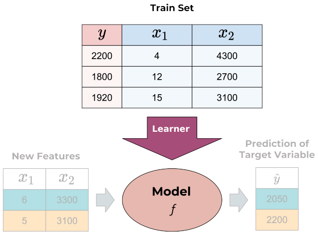
Introduction to Machine Learning

ML-Basics

Learner



Learning goals

- Know formal definition of learner
- Understand that a learner receives training data and outputs the best model from $\mathcal{H}$

SUPERVISED LEARNING EXAMPLE

- Imagine we want to investigate how working conditions affect productivity of employees
- It is a **regression** task since the target *productivity* is continuous
- We collect data about worked minutes per week (*productivity*), how many people work in the same office as the employee in question, and the employee's salary

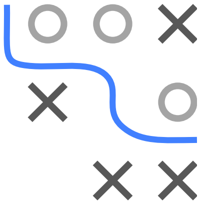
Features x		Target y
People in Office (Feature 1) x_1	Salary (Feature 2) x_2	Worked Minutes Week (Target Variable)
4	4300 €	2220
12	2700 €	1800
5	3100 €	1920

$n = 3$

$p = 2$

$x_1^{(2)}$ $x_2^{(1)}$ $y^{(3)}$

SUPERVISED LEARNING EXAMPLE



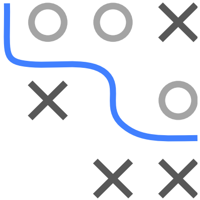
How could we construct a model from these data?

We could investigate the data manually and come up with a simple, hand-crafted rule such as:

- The baseline productivity of an employee with salary 3000 and 7 people in the office is 1850 minutes
- A decrease of 1 person in the office increases productivity by 30
- An increase of the salary by 100 increases productivity by 10

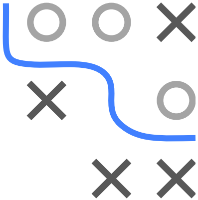
$\Rightarrow$ Obviously, this is neither feasible nor leads to a good model.

IDEA OF SUPERVISED LEARNING

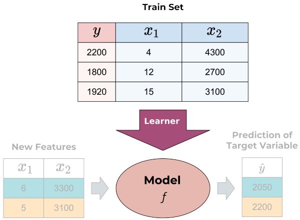


- **Goal:** Identify the functional relationship that maps features to target
- **Supervised** learning means we use *labeled* data to learn model *f*
- Later, we use model *f* to predict *y* for new *unlabeled* data

LEARNER DEFINITION



- Algorithm for finding f is called **learner** / **learning algorithm** / **inducer**
- The learner is our means of picking the best element from the hypothesis space $\mathcal{H}$ for given training data
- Formally it maps training data $\mathcal{D} \in \mathbb{D}$ (plus a vector of **hyperparameter** control settings $\boldsymbol{\lambda} \in \Lambda$) to a model:



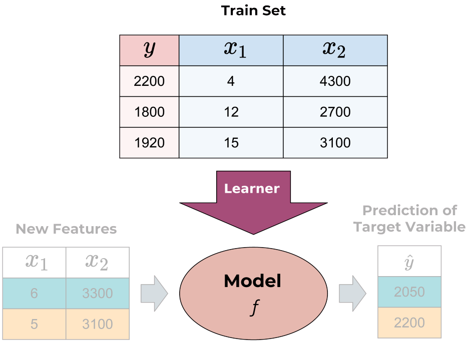
$$\mathcal{I} : \mathbb{D} \times \Lambda \rightarrow \mathcal{H}$$

- Practically, we often construct a mapping $\mathcal{I} : \mathbb{D} \times \Lambda \rightarrow \Theta$

LEARNER DEFINITION

In pseudo-code:

- Learner gets a hypothesis space of parametrized functions $\mathcal{H}$
- User passes data set $\mathcal{D}_{\text{train}}$ and control settings $\boldsymbol{\lambda}$
- Learner sets parameters such that model fits data best
- Optimal parameters $\hat{\theta}$ or function $\hat{f}$ is returned for later usage



y	x_1	x_2
2200	4	4300
1800	12	2700
1920	15	3100

x_1	x_2
6	3300
5	3100

$\hat{y}$
2050
2200