Introduction to Machine Learning

Introduction: Supervised Learning & Learning Tasks

compstat-lmu.github.io/lecture_i2ml

IDEA OF SUPERVISED LEARNING

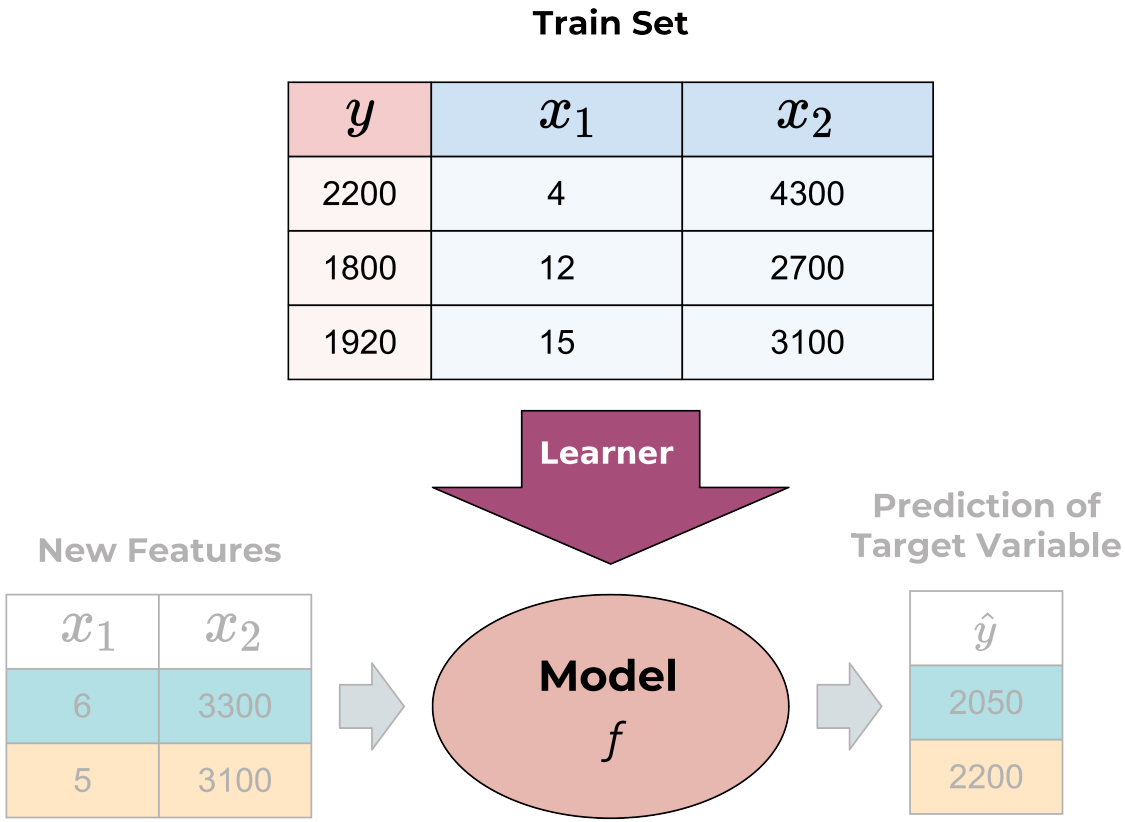
- **Goal:** Identify the fundamental functional relation in the data that maps an object's features to the target.
- Ideally, we would have full knowledge about the data-generating process and thus be able to specify this mapping function precisely.
- However, since this is basically impossible, we must try to **learn** the mapping function: for objects exhibiting certain patterns or properties, certain outcomes are much more likely.
  $\rightarrow$ We call such an assumed mapping a **model** f.

IDEA OF SUPERVISED LEARNING

- **Supervised** learning means we make use of *labeled* data, i.e., observations for which we already know the target outcome.
- We try to construct f automatically from an example set of such labeled objects.
  $\rightarrow$ The algorithm for finding f is called **learner**.
- Using the thus learned model, we can make **predictions** based on the features of our data.
- Knowing the "truth" allows us to test how well we have grasped the nature of the underlying mapping: we just need to compare our predictions to the actually observed values.

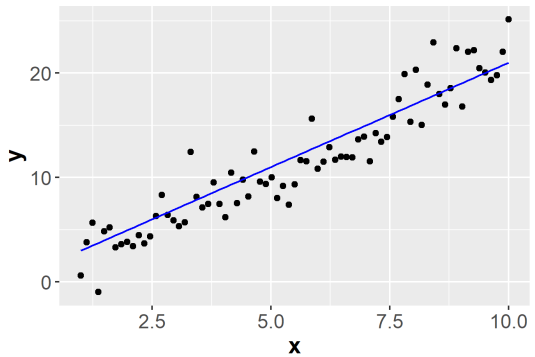
IDEA OF SUPERVISED LEARNING

- Ultimately, we will use our model to compute predictions for **new** data whose target values are unknown.

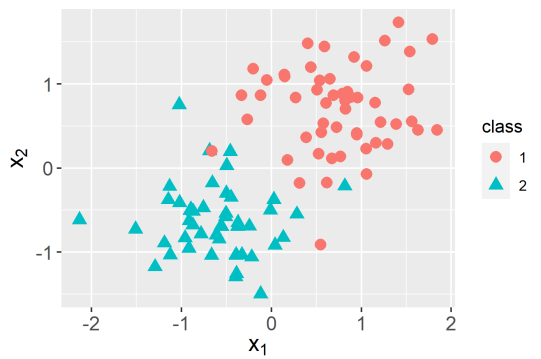


TASKS IN SUPERVISED LEARNING

- In general, supervised learning comes in two flavors we call **tasks**:
  - **Regression**: Given features **x**, predict corresponding output from $\mathcal{Y} \in \mathbb{R}^m$.

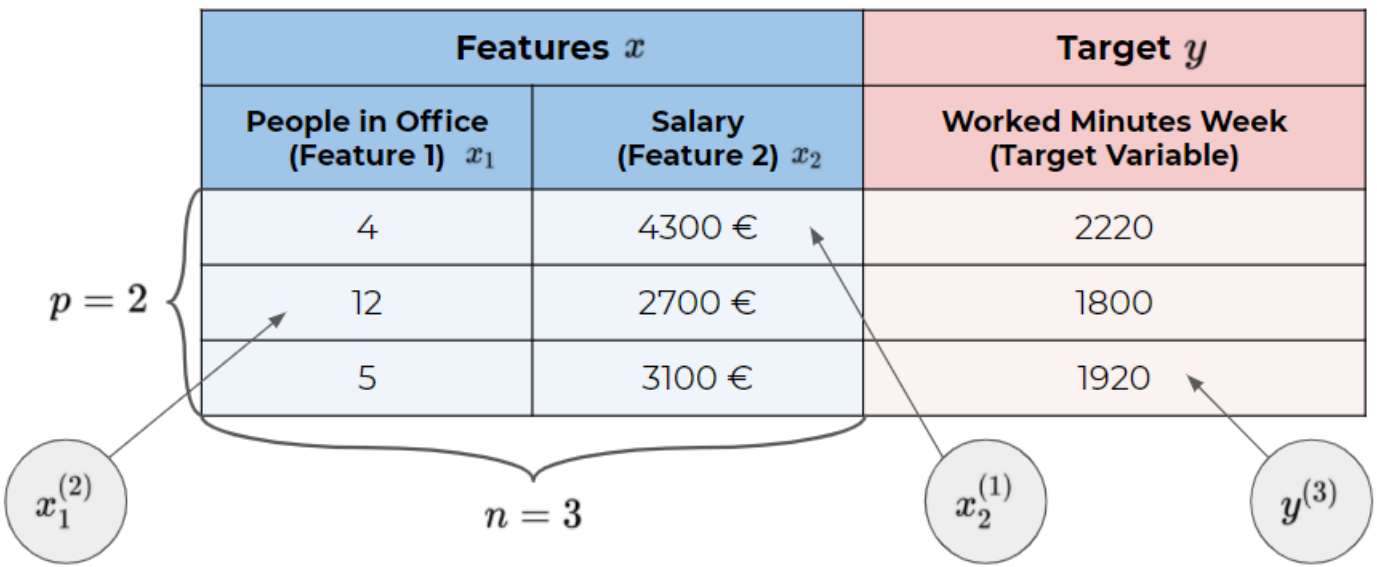


  - **Classification**: Assign an observation with features **x** to one class of a finite set of classes $\mathcal{Y} = \{C_1, ..., C_g\}, g \geq 2$ (details later).



REGRESSION TASKS: EXAMPLE

Imagine you want to investigate how salary and workplace conditions (*features*) affect productivity of employees (*target*) – a standard **regression** task. Therefore, you collect data about their worked minutes per week (productivity), how many people work in the same office as the employees in question, and the employees' salary.



REGRESSION TASKS: EXAMPLE

- For our observed data we know which outcome is produced.
- For new employees can only observe the features but not the target.

MORE REGRESSION TASKS

1. **House Prices**

   Aim: Predict the price for a house in a certain area

Features x				Target y
square footage of the house	number of bedrooms	swimming pool (yes/no)	...	house price in US$
1,180	3	0	...	221,900
2,570	3	1	...	538,000
770	2	0	...	180,000
1,960	4	1	...	604,000

MORE REGRESSION TASKS

2. **Length-of-stay in a hospital**

   Aim: Predict the number of days a single patient has to stay in hospital at the time of admission

Features x					Target y
diagnosis category	**admission type**	**gender**	**age**	**...**	**Length-of-stay in the hospital in days**
heart disease	elective	male	75	...	4.6
injury	emergency	male	22	...	2.6
psychosis	newborn	female	0	...	8
pneumonia	urgent	female	67	...	5.5

CLASSIFICATION TASKS: EXAMPLE

Risk category in life insurance

Aim: Predict one of five risk categories (**classification**) for a life insurance customer to determine the insurance premium

Features x				Target y
job type	**age**	**smoker**	**...**	**risk group**
carpenter	34	1	...	3
stuntman	25	0	...	5
student	23	0	...	1
white-collar worker	39	0	...	2