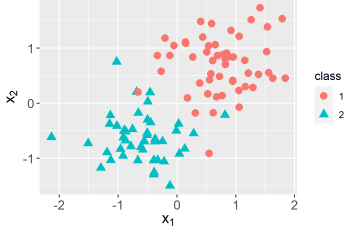
Introduction to Machine Learning

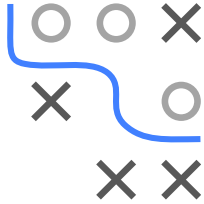
ML-Basics: Supervised Tasks



Learning goals

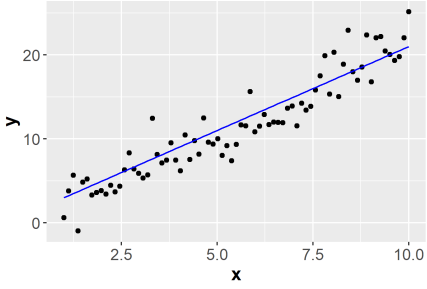
- Know definition and examples of supervised tasks
- Understand the difference between regression and classification

TASKS: REGRESSION VS CLASSIFICATION

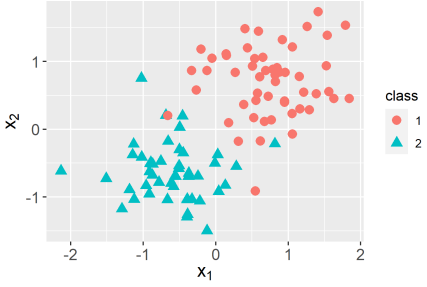


- Supervised tasks are data situations where learning the functional relationship between inputs (features) and output (target) is useful.
- The two most basic tasks are regression and classification, depending on whether the target is numerical or categorical.

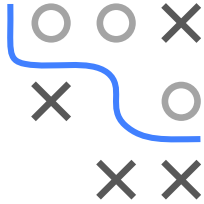
Regression: Our observed labels come from $\mathcal{Y} \subseteq \mathbb{R}$.



Classification: Observations are categorized: $y \in \mathcal{Y} = \{C_1, ..., C_g\}$.



PREDICT VS. EXPLAIN



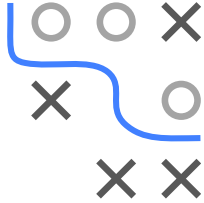
We can distinguish two main reasons to learn this relationship:

- **Learning to predict**. In such a case we potentially do not care how our model is structured or whether we can understand it.
  Example: predicting how a stock price will develop.
  Simply being able to use the predictor on new data is of direct benefit to us.
- **Learning to explain**. Here, our model is only a means to a better understanding of the inherent relationship in the data.
  Example: understanding which risk factors influence the probability to get a certain disease. We might not use the learned model on new observations, but rather discuss its implications, in a scientific or social context.

While ML was traditionally more interested in the former, classical statistics addressed the latter. In many tasks nowadays both are relevant – to different degrees.

REGRESSION EXAMPLE: HOUSE PRICES

Predict the price for a house in a certain area

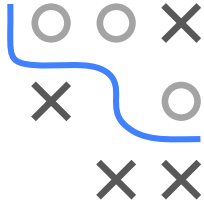


Features x				Target y
square footage of the house	number of bedrooms	swimming pool (yes/no)	...	house price in US$
1,180	3	0	...	221,900
2,570	3	1	...	538,000
770	2	0	...	180,000
1,960	4	1	...	604,000

Probably *learn to explain*. We might want to understand what influences a house price most. But maybe we are also looking for underpriced houses and the predictor is of direct use, too.

REGRESSION EXAMPLE: LENGTH-OF-STAY

Predict days a patient has to stay in hospital at time of admission

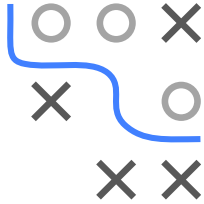


Features x					Target y
diagnosis category	admission type	gender	age	...	Length-of-stay in the hospital in days
heart disease	elective	male	75	...	4.6
injury	emergency	male	22	...	2.6
psychosis	newborn	female	0	...	8
pneumonia	urgent	female	67	...	5.5

Can be *learn to explain*, but *learn to predict* would help a hospital's planning immensely.

CLASSIFICATION EXAMPLE: RISK CATEGORY

Predict one of five risk categories 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

Probably *learn to predict*, but the company might be required to explain its predictions to its customers.