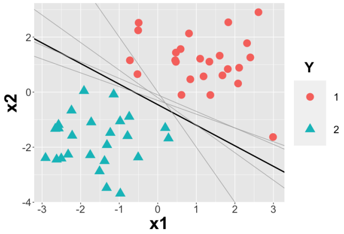
Introduction to Machine Learning

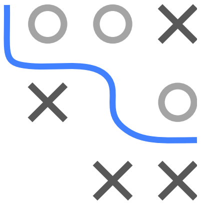
ML-Basics
In a Nutshell



Learning goals

- Understand fundamental goal of supervised machine learning
- Know concepts of task, model, parameter, learner, loss function, and empirical risk minimization

WHAT IS ML?

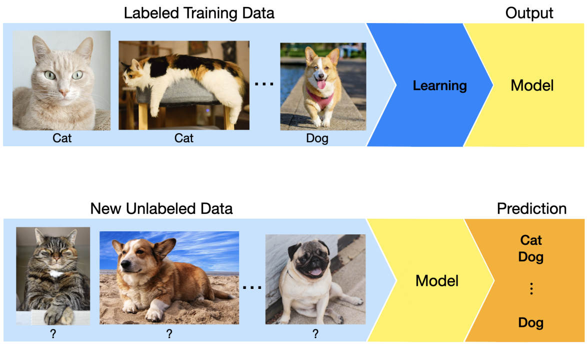


"A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E."
Tom Mitchell, Carnegie Mellon University, 1998

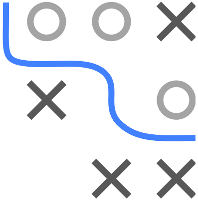
$\Rightarrow$ 99 % of this lecture is about **supervised learning**:

Training

Prediction

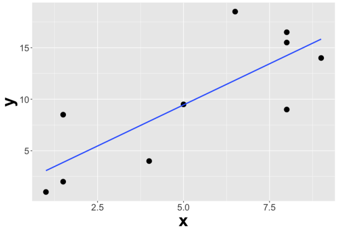


TASKS

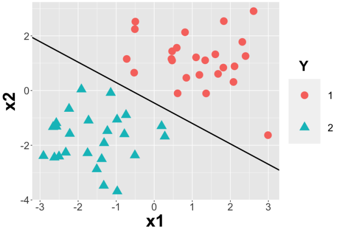


- Supervised tasks are labeled data situations where the goal is to learn the functional relationship between inputs (features) and output (target)
- We distinguish between **regression** and **classification** tasks, depending on whether the target is **numerical** or **categorical**

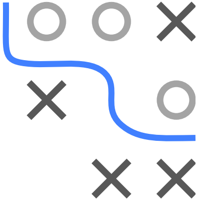
Regression: Target is **numerical**, e.g., predict days a patient has to stay in hospital



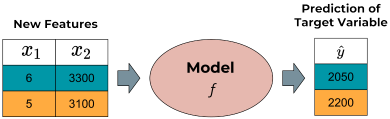
Classification: Target is **categorical**, e.g., predict one of two risk categories for a life insurance customer



MODELS AND PARAMETERS

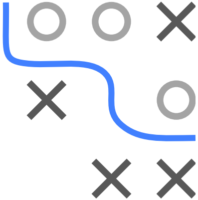


- A model is a function that maps features to predicted targets



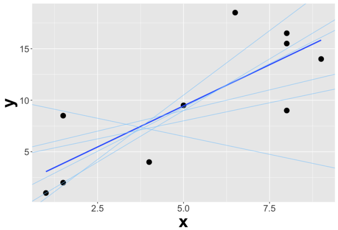
- For finding the model that describes the relation between features and target best, one needs to restrict the set of all possible functions
- This restricted set of functions is called **hypothesis space**. E.g., one could consider only simple linear functions as hypothesis space
- Functions are fully determined by parameters. E.g., in the case of linear functions, $y = \theta_0 + \theta_1 x$, the parameters θ_0 (intercept) and θ_1 (slope) determine the relationship between y and x
- Finding the optimal model means finding the optimal set of parameters

LEARNER

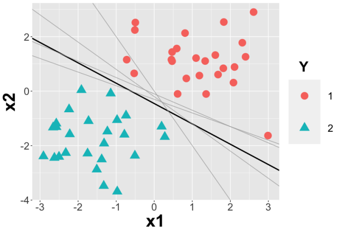


- Learns automatically the relation between features and target – given a set of training data
- Learner picks the best element of the **hypothesis space**, i.e., the function that fits the training data best

Regression:

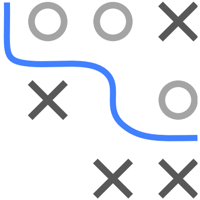


Classification:



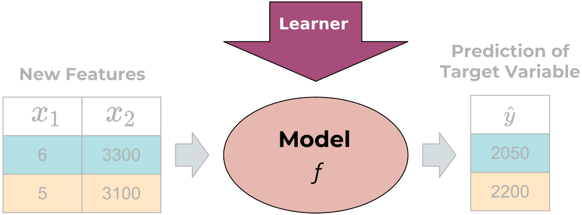
LEARNER / 2

- Learner uses labeled training data to learn a model f. This model is applied to new data for predicting the target variable



Train Set

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



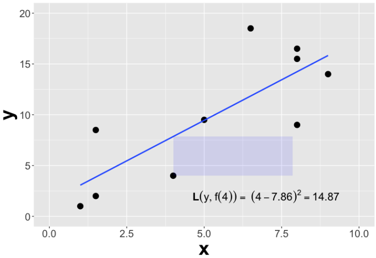
x_1	x_2
6	3300
5	3100

$\hat{y}$
2050
2200

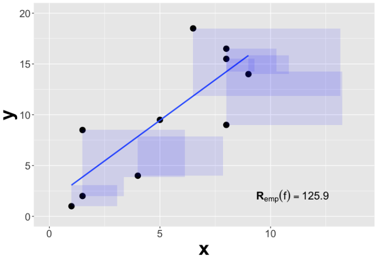
LOSS AND RISK MINIMIZATION

- Loss: Measured pointwise for each observation, e.g., L_2-loss
$$L\left(y, f(\mathbf{x})\right) = \left(y - f(\mathbf{x})\right)^2$$
- Risk: Measured for entire model. Sums up pointwise losses.
$$\mathcal{R}_{\text{emp}}(f) = \sum_{i=1}^{n} L\left(y^{(i)}, f\left(\mathbf{x}^{(i)}\right)\right)$$

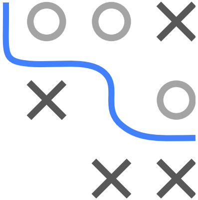
Squared **loss** of one **observation**.



Empirical **risk** of entire **model**



EMPIRICAL RISK MINIMIZATION



- The risk surface visualizes the empirical risk for all possible parameter values of the parameter vector $\boldsymbol{\theta}$
- Minimizing the empirical risk is usually done by numerical optimization

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta} \in \Theta} \mathcal{R}_{\text{emp}}(\boldsymbol{\theta}).$$