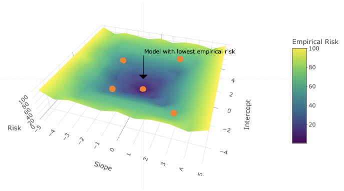
Introduction to Machine Learning

ML-Basics

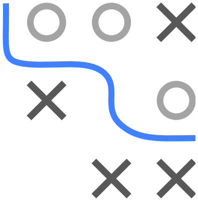
In a Nutshell



Learning goals

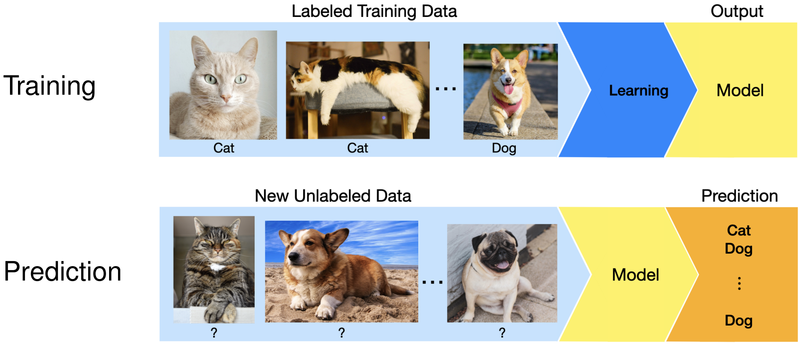
- Understand fundamental goal of supervised machine learning
- Know concepts of data, task, model, parameter, learner, loss function, and empirical risk minimization

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

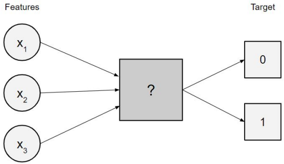
$\Rightarrow$ 99 % of this lecture is about **supervised learning**:



DATA IN SL

Measurements on different aspects of *n* objects:

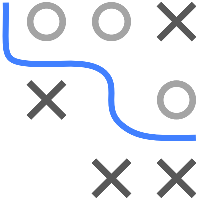
- **Target**: output variable / goal of prediction
- **Features**: properties that describe an object



Features x				Target y
Sepal.Length	**Sepal.Width**	**Petal.Length**	**Petal.Width**	**Species**
4.3	3.0	1.1	0.1	setosa
5.0	3.3	1.4	0.2	setosa
7.7	3.8	6.7	2.2	virginica
5.5	2.5	4.0	1.3	versicolor

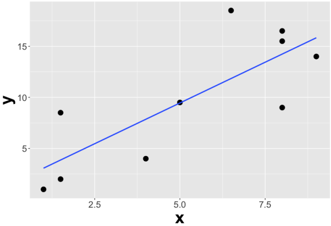
$$\mathcal{D} = \left(\left(\mathbf{x}^{(1)}, y^{(1)}\right), \ldots, \left(\mathbf{x}^{(n)}, y^{(n)}\right)\right) \in (\mathcal{X} \times \mathcal{Y})^n$$

TASKS

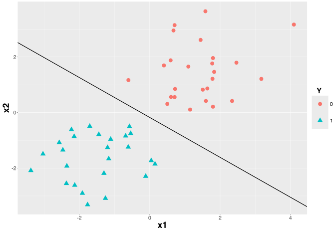


- Supervised tasks are labeled data situations with the goal of learning the functional relationship between features and target
- We distinguish between **regression** and **classification** tasks, depending on whether the target is **numerical** or **categorical**

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



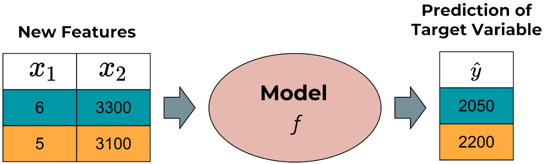
Classification: Target is **categorical**, e.g., predict one of two risk categories for a life insurance customer



MODELS AND PARAMETERS

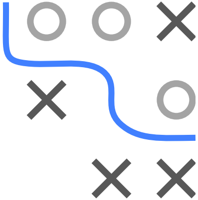
- Maps features to predicted targets

$$f_\theta : \mathcal{X} \to \mathbb{R}^g$$



- We consider a restricted set of parametrized functions of a certain form, the **hypothesis space**, e.g., simple linear functions
- Models 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

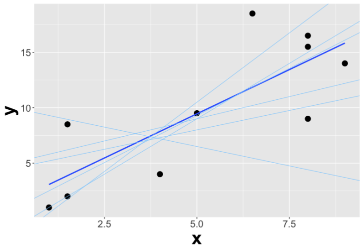
LEARNER



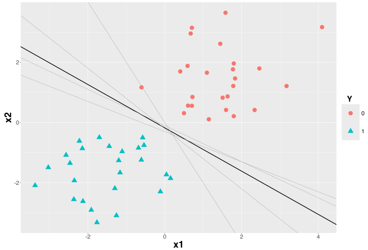
- Learns automatically the relation between features and target – given a set of training data
- Learner picks the best element of the **hypothesis space**, i.e., the function that fits the training data best:

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

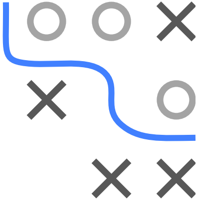
Regression



Classification



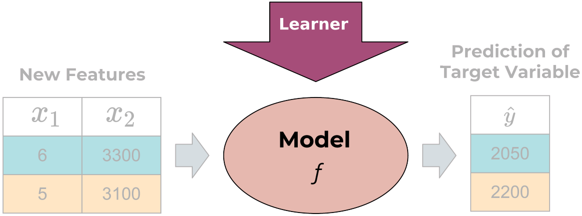
LEARNER



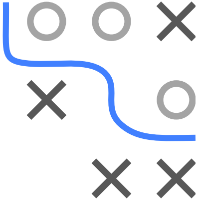
- Learner uses labeled training data to learn a model f. This model is applied to new unlabeled data for predicting the target variable

Train Set

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



LOSS AND RISK MINIMIZATION



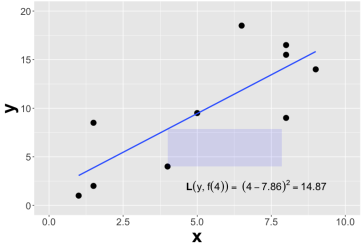
- Loss: Measured pointwise for each observation, e.g., L_2-loss

$$L\left(y, f(\mathbf{x})\right) = \left(y - f(\mathbf{x})\right)^2$$

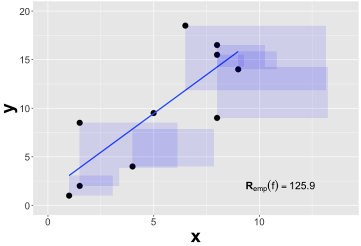
- Risk: Measured for entire model. Sums up pointwise losses

$$\mathcal{R}_{\text{emp}}(f) = \sum_{i=1}^{n} L\left(y^{(i)}, f\left(\mathbf{x}^{(i)}\right)\right)$$

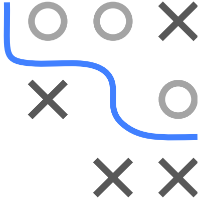
Squared **loss** of one **observation**



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 often done by numerical optimization

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