Introduction to Machine Learning

Introduction: Models & Parameters

compstat-lmu.github.io/lecture_i2ml

WHAT IS A MODEL?

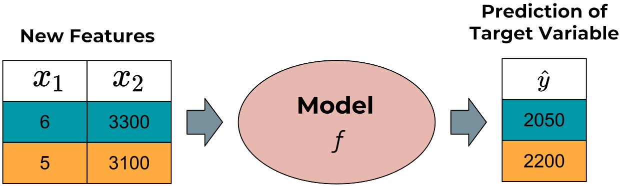
- A **model** (or **hypothesis**)

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

  is a function that maps feature vectors to predicted target values.
- As such, it is (the attempt at) a **formal representation** of the observed data.
- In conventional regression we will have $g = 1$, for classification see later.
- f is meant to capture intrinsic patterns of the data, the underlying assumption being that these patterns hold true not only for the observed sample but for all data drawn from $\mathbb{P}_{xy}$.

WHAT IS A MODEL?

- It is easily conceivable how models can range from super simple to incredibly complex.
- The ultimate goal is to *generalize* the learned model to new data (we already know the outcome for our training data), with as little error as possible.
- This suggests that we might be interested in a certain simplifying property: a model is expected to perform complexity reduction.
  $\rightarrow$ It needs to be scalable and extendable to new data situations.



HYPOTHESIS SPACES

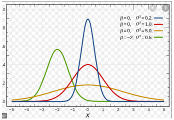
- We have already seen that machine learning typically requires constraining f to a certain class of functions.
- Otherwise, the task of finding a “good” model among all the available ones is basically impossible to solve.
- The set of functions defining a specific model class is called a **hypothesis space** $\mathcal{H}$.
- For example, the set of all linear functions through $(0|0)$

$$\mathcal{H} = \{f : f(\mathbf{x}) = c\mathbf{x}, c \in \mathbb{R}\}$$

forms a (rather simple) hypothesis space.

PARAMETERS OF A MODEL

- Within one hypothesis space, models are "alike" in a sense: they all share a common structure that makes up the condition in defining $\mathcal{H}$.
  $\rightarrow$ E.g., all Gaussian density functions are of a bell-like shape.
- Of all models in a class it is the choice of **parameter** values that singles out a specific representant $f \in \mathcal{H}$.
- Parameters are our means of configuration: once set, our model is fully determined.
  $\rightarrow$ Gaussians are solely determined by mean μ and variance σ^2.
- Parameters are the instrument to tailor the general hypotheses to our data.



PARAMETERS OF A MODEL

- We usually subsume all parameters in a **parameter vector** $\boldsymbol{\theta} = (\theta_1, \theta_2, ...)$ from a **parameter space** Θ.
- $\boldsymbol{\theta}$ might be one-dimensional or comprise thousands of parameters, depending on the complexity of our model.
- $\boldsymbol{\theta}$ is what we try to learn during training: finding a "good" model boils down to finding a suitable combination of parameters.
- We will see in the next chapter how the "goodness" of a model can be determined.