Introduction to Machine Learning

Introduction: Models & Parameters

compstat-lmu.github.io/lecture_i2ml

WHAT IS A MODEL?

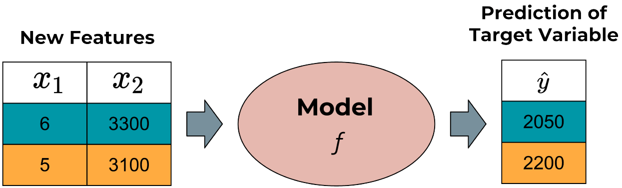
- A **model** (or **hypothesis**)

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

  is a function that maps feature vectors to predicted target values.
- f is meant to capture intrinsic patterns of the data, the underlying assumption being that these patterns hold true for *all* data drawn from $\mathbb{P}_{xy}$.
- Loosely speaking: if f is fed a set of features, it will output the target corresponding to these feature values under our hypothesis.

In conventional regression we will have $g = 1$; for classification g equals the number of classes, and output vectors are scores or class probabilities (details later).

WHAT IS A MODEL?



- It is easily conceivable how models can range from super simple (e.g., tree stumps) to reasonably complex (e.g., variational autoencoders), and how there is an infinite number of them.
- In fact, machine learning requires **constraining** f to a certain type of functions.

HYPOTHESIS SPACES

- Without restrictions on the functional family, the task of finding a "good" model among all the available ones is impossible to solve.
- This means: we have to determine the class of our model *a priori*, thereby narrowing down our options considerably.
- The set of functions defining a specific model class is called a **hypothesis space** $\mathcal{H}$:

$$\mathcal{H} = \{f : f \text{ belongs to a certain functional family}\}$$

HYPOTHESIS SPACES

- **Example 1:** Hypothesis space of univariate linear functions

$$\mathcal{H} = \{f : f(\mathbf{x}) = \boldsymbol{\theta}^T \mathbf{x}, \boldsymbol{\theta} \in \mathbb{R}^p\}$$

FIGURE

- **Example 2:** Hypothesis space of bivariate quadratic functions

$$\mathcal{H} = \{f : f(\mathbf{x}) = \boldsymbol{\theta}^T \mathbf{x}, \boldsymbol{\theta} \in \mathbb{R}^p\}$$

FIGURE

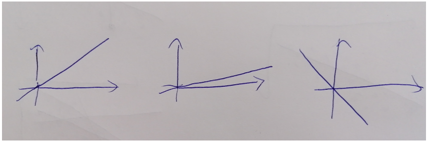
- **Example 3:** Hypothesis space of radial basis function networks

$$\mathcal{H} = \{f : f(\mathbf{x}) = \boldsymbol{\theta}^T \mathbf{x}, \boldsymbol{\theta} \in \mathbb{R}^p\}$$

FIGURE

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 $\mathcal{H}$.
  $\rightarrow$ In the above example, all models are straight lines passing through the origin.
- 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.
  $\rightarrow$ Origin-crossing lines are solely determined by their slope, c.
- 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.

PARAMETERS, STATISTICS AND SUPERVISED ML

- Supervised ML additionally assumes that f is of a certain “form” or comes from a certain **class of functions**.
  This is necessary to make the problem of automatically finding a “good” model feasible at all.
- The specific behavior of a mapping from this class can then be described by **parameters** that define its shape.
- Statistics, too, studies how to learn such functions (or, rather: their parameters) from example data and how to perform inference on them and interpret the results.
- For historical reasons though, statistics is mostly focused on fairly simple classes of mappings, like (generalized) linear models.
- Supervised ML also includes more complex kinds of mappings that can typically deal with more complicated and high-dimensional inputs.