Exercise Sheet 4

Due: 05.12.2018, 09:00

Download the notebook **poly_fit.ipynb** from ISIS.

Exercise 4.1

Let $x_1, x_2, \dots, x_m$ be drawn independently from distribution P with mean μ. Show that the sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is an unbiased estimator of the population mean μ, that is $E(\bar{x}) = \mu$.

Exercise 4.2 (solving one of a) and b) is enough to get a point)

a) Give an example of a hypothesis space $\mathcal{H}$, where the expected hypothesis $h^*(x) = \mathbb{E}_{\mathcal{D}}[h_{\mathcal{D}}(x)]$ is always contained in $\mathcal{H}$.
b) Give an example of a hypothesis space $\mathcal{H}$, where the expected hypothesis $h^*(x) = \mathbb{E}_{\mathcal{D}}[h_{\mathcal{D}}(x)]$ is not guaranteed to be contained in $\mathcal{H}$. That means, there is a probability distribution on input-output examples $P(x, y)$ which yields $h^*(x) \notin \mathcal{H}$. Hint: It is enough to argue that there may be such a $P(x, y)$, it is not necessary to specify such a $P(x, y)$.

Exercise 4.3

The bias variance decomposition revealed that we want to find a learner that simultaneously has low bias and low variance. Explain why there is a tradeoff between minimizing bias and minimizing variance and how the tradeoff can be controlled.

Exercise 4.4

The goal of this exercise is to empirically investigate the relation between bias, variance and generalization error in dependence of the model complexity.
Assume that we observe data $y = f(x) + \varepsilon$, where ε is normally distributed with zero mean and variance σ^2 and

$$f: [-1,1] \to \mathbb{R}, \quad x \mapsto \tanh(10x).$$

Use Scikit-learn (poly_fit.ipynb contains a demonstration) to fit polynomials of order 9 to the observed data for varying L_2-regularization parameter λ. For each standard deviation $\sigma \in \{0, 0.5\}$ perform the bias-variance decomposition as follows:

- Generate training datasets of fixed size $m = 10$ where input values x are drawn independently from the uniform distribution on $[-1,1]$ and output values are given by $y = f(x) + \varepsilon$.
- Use empirical estimates of bias, variance, bias+variance and the generalization error of the learner and plot them in dependence of the regularization parameter λ. Choose a few suitable regularization parameters for this.

Discuss the results based on the following questions:

- Which regularization parameter is the best?
- Determine for this regularization parameter the deviation of Bias+Variance to the generalization error. Explain the value of the deviation.
- Can we estimate the generalization error in practice using the bias-variance decomposition? Why / why not?