

Homework 2

Due on 2nd November

Problem 2.24

Consider a simplified learning scenario. Assume that the input dimension is one. Assume that the input variable x is uniformly distributed in the interval $[-1, 1]$. The data set consists of 2 points $\{x_1, x_2\}$ and assume that the target function is $f(x) = x^2$. Thus, the full data set is $\mathcal{D} = \{(x_1, x_1^2), (x_2, x_2^2)\}$. The learning algorithm returns the line fitting these two points as g (\mathcal{H} consists of functions of the form $h(x) = ax + b$). We are interested in the test performance (E_{out}) of our learning system with respect to the squared error measure, the bias and the variance.

- (a) Give the analytic expression for the average function $\bar{g}(x)$.
- (b) Describe an experiment that you could run to determine (numerically) $\bar{g}(x)$, E_{out} , **bias**, and **var**.
- (c) Run your experiment and report the results. Compare E_{out} with **bias+var**. Provide a plot of your $\bar{g}(x)$ and $f(x)$ (on the same plot).
- (d) Compute analytically what E_{out} , **bias**, and **var** should be.