Elements of Machine Learning
Exercise Sheet 1
Winter Term 2023/2024

William LaCroix - wila00001@stud.uni-saarland.de - 7038732
Philipp Hawlitschek - phha00002@stud.unisaarland.de - 7043167

Problem 4 (T, 6 points). Bias and Variance.

Consider the bias and variance of a linear model f.

1. [1pts] Explain in concise terms the meaning of bias and variance in the context of linear regression. What is the relationship between them?

Bias in regression is how flexible the model is. A highly flexible model may fit the data, but not necessarily the underlying distribution, whereas an inflexible model introduces bias in an attempt to roughly approximate highly irregular data.

Variance is the difference in a given dataset from the underlying distribution. Models with high variance may approximate one sample, while completely failing to predict the data from another.

2. [2pts] Consider the following equation,

$$Var\left(\hat{f}\left(x_0\right)\right) + [Bias\left(\hat{f}\left(x_0\right)\right)]^2 = \mathbb{E}[\left(f\left(x_0\right) - \hat{f}\left(x_0\right)\right)^2].$$

Explain the meaning of each term and show that the above holds.

$Var\left(\hat{f}\left(x_0\right)\right)$ the variance in our model at a given data point, x_0, this is a measure of the difference in model prediction across datasets.

$[Bias\left(\hat{f}\left(x_0\right)\right)]^2$ is the model bias, representing how well the model fits the underlying function. If the model is not flexible enough, it will not fit the underlying distribution well, leading to a high bias term.

$\mathbb{E}[\left(f\left(x_0\right) - \hat{f}\left(x_0\right)\right)^2]$ is the mean square error, being the average squared difference between our model's prediction and the observed data.

$Var\left(\hat{f}\left(x_0\right)\right) = \mathbb{E}[\hat{f}\left(x_0\right) - \mathbb{E}\left(\hat{f}\left(x_0\right)\right)]^2 = \mathbb{E}[\hat{f}\left(x_0\right)]^2 - \mathbb{E}[\left(\hat{f}\left(x_0\right)\right)]^2$
$[Bias\left(\hat{f}\left(x_0\right)\right)]^2 = [\mathbb{E}\left(\hat{f}\left(x_0\right) - f\left(x_0\right)\right)]^2$

$$\begin{aligned}
\mathbb{E}[\left(f\left(x_0\right) - \hat{f}\left(x_0\right)\right)^2] &= Var\left(\hat{f}\left(x_0\right)\right) + [Bias\left(\hat{f}\left(x_0\right)\right)]^2 \\
&= \mathbb{E}[\hat{f}\left(x_0\right)]^2 - \mathbb{E}[\left(\hat{f}\left(x_0\right)\right)]^2 + [\mathbb{E}\left(\hat{f}\left(x_0\right) - f\left(x_0\right)\right)]^2 \\
&= \mathbb{E}[\hat{f}\left(x_0\right)]^2 - \mathbb{E}[\left(\hat{f}\left(x_0\right)\right)]^2 + \mathbb{E}[\left(\hat{f}\left(x_0\right)^2] - \mathbb{E}[2\hat{f}\left(x_0\right)f\left(x_0\right)\right)] + \mathbb{E}[f\left(x_0\right)^2] \\
&= \mathbb{E}[\hat{f}\left(x_0\right)]^2 - \mathbb{E}[2\hat{f}\left(x_0\right)f\left(x_0\right)] + \mathbb{E}[f\left(x_0\right)^2] \\
&= \mathbb{E}[\hat{f}\left(x_0\right)^2 - 2\hat{f}\left(x_0\right)f\left(x_0\right) + f\left(x_0\right)^2] \\
&= \mathbb{E}[\left(f\left(x_0\right) - \hat{f}\left(x_0\right)\right)^2]
\end{aligned}$$

3. [2pts] How is $\mathbb{E}[\left(f\left(x_0\right) - \hat{f}\left(x_0\right)\right)^2]$ related to the expected test MSE, $\mathbb{E}[\left(y_0 - \hat{f}\left(x_0\right)\right)^2]$? Consider the difference of these quantities and explain its meaning.

$\mathbb{E}[\left(f\left(x_0\right) - \hat{f}\left(x_0\right)\right)^2]$ here is the expected squared difference between predicted values, and the values observed in training at a given data point, x_0. $\mathbb{E}[\left(y_0 - \hat{f}\left(x_0\right)\right)^2]$ on the other hand is the difference between the true value (y_0) and the predicted value of $f\left(x_0\right)$ using the model trained on a separate data set.

4. [1pts] State whether the following statement is true or false and explain why.

"The Gauss-Markov theorem states that the least-squares estimates $\hat{\beta}$ have the smallest variance among all linear estimates. Since the least-squares estimates $\hat{\beta}$ are unbiased, this means that biased estimators will always have a larger variance than $\hat{\beta}$."

The statement is false. GMT states that **among unbiased estimators**, OLS has the lowest variance. It makes no claims about biased estimators.