Elements of Machine Learning
Assignment 1

William LaCroix (7038732)
wila00001@stud.uni-saarland.de

Problem 3 (T, 6 Points). **Gauss-Markov Theorem**

1. [2pts] Explain in your own words what the Gauss-Markov theorem is and why it is important.

The Gauss-Markov Theorem (abbreviated GMT from now on) is a framework for minimizing least-squares coefficients as estimates of linear functions of X. This is useful, because these coefficients display the lowest sampling variance among all unbiased linear estimates.

2. [3pts] Explain in your own words and write down the formal math formulation of the three *Gauss-Markov error term assumptions.*

Assumption 1: error is mean-zero – $\forall_i : E[\epsilon_i] = 0$

The average value of the error variables is centered at zero, so while observed values will be both above and below the true function, the mean lies on that function.

Assumption 2: errors are homoscedastic – $\forall_i : Var(\epsilon_i) = \sigma^2$

The variance in the error term stays constant and does not depend upon the response variable.

Assumption 3: error terms are uncorrelated – $\forall_i | i \neq j : Cov(\epsilon_i, \epsilon_j) = 0$

The error terms do not depend on each other. This would lead to larger than predicted error values, since the errors will not be constant between observations.

3. [1pts] Is it also the *best* (in terms of test error) linear unbiased estimate (argue with the bias-variance trade-off)?

It *is* the best, since it has the lowest sample variance among unbiased linear estimates. Some models may be able to achieve smaller variance by setting certain predictor coefficients to zero or becoming highly non-linear, but both of these approaches improve variance by increasing model bias. With Ordinary Least Squares estimators, the error of the estimate does not depend on the predictor or other errors, and given mean-zero error variance, allows a functional estimate to be derived where the total error variance approaches the irreducible error as the model approaches the true function.