NNTI Assignment 04

Name 1: Md Mobashir Rahman
Student ID 1: 7059086
Email 1: mdra00001@stud.uni-saarland.de

Name 2: Ratnadeep Chakraborty
Student ID 2: 7022859
Email 2: rach00002@stud.uni-saarland.de

4.1.A

Bias: Bias measures the error that is caused by the estimator's inability to capture the true patterns in the data. It leads to underfitting.

Variance: Variance measures the error caused by the fluctuation in the dataset. High variance indicates overfitting of the data and high complexity of the model.

Bias-Variance Trade-off:

Bias decreases as model complexity increases; however, variance increases when model complexity increases.

A good estimator balances bias and variance, and minimizes error.

Overfitting and Underfitting

Overfitting causes low bias but high variance, and underfitting causes high bias but low variance.

4.1.B

Increase the width of hidden layers:

$$\text{Bias} \downarrow, \quad \text{Variance} \uparrow$$

Wider neural networks can learn patterns in data very well, often causing overfitting, which results in low bias and high variance.

Increasing polynomial degree n:

$$\text{Bias} \downarrow, \quad \text{Variance} \uparrow$$

Higher polynomial degree overfits the data.

Training the model on less data:

$$\text{Bias:}\quad \text{NEI} \quad \text{Variance:} \uparrow$$

Lack of data can cause overfitting and instability, resulting in a high variance, but bias remains unchanged.

Adding regularization:

$$\text{Bias:} \uparrow \quad \text{Variance:} \downarrow$$

Regularization penalizes the model when it is too complex. It increases bias and decreases variance by preventing the model from overfitting.

Removing some non-support vectors:

$$\text{Bias:} \quad - \quad \text{Variance:} \downarrow$$

Non-support vectors don't change the margins, so bias does not change; however, variance decreases.

Decreasing hypotheses space:

$$\text{Bias:} \uparrow \quad \text{Variance:} \downarrow$$

Reducing hypotheses space makes the model simpler and causes high bias and low variance.

4.1.C

By definition,

$$\text{MSE}(y_{\text{test}}, \hat{f}_n) = E\left[(y_{\text{test}} - \hat{f}_n)^2\right]$$

Given,

$$y_{\text{test}} = f(x) + \epsilon,$$

$$E[\epsilon] = 0,$$

$$\text{Var}(\epsilon) = \tau^2,$$

Substituting y_{test} in the MSE equation,

$$\text{MSE}(y_{\text{test}}, \hat{f}_n) = E\left[(f(x) + \epsilon - \hat{f}_n)^2\right]$$

Expanding $(f(x) + \epsilon - \hat{f}_n)^2$,

$$(f(x) - \hat{f}_n + \epsilon)^2$$

$$= (f(x) - \hat{f}_n)^2 + 2\epsilon(f(x) - \hat{f}_n) + \epsilon^2$$

Taking the expectation,

$$E\left[(f(x) - \hat{f}_n)^2\right] + E\left[2\epsilon \cdot (f(x) - \hat{f}_n)\right] + E[\epsilon^2]$$

Given $\mathrm{E}[\epsilon] = 0$ and $E[\epsilon^2] = \mathrm{Var}(\epsilon) = \tau^2$,

$$\mathrm{MSE} = E\left[(f(x) - \hat{f}_n)^2\right] + \tau^2$$

Decomposing $\mathrm{E}\left[(f(x) - \hat{f}_n)^2\right]$ by adding and subtracting the expected prediction $E[\hat{f}_n], E\left[(f(x) - \hat{f}_n)^2\right]$ $E\left[(f(x) - E[\hat{f}_n] + E[\hat{f}_n] - \hat{f}_n)^2\right]$

$$= E\left[(f(x) - E[\hat{f}_n])^2 + E[\hat{f}_n] - \hat{f}_n)^2\right]$$

$$= E\left[(f(x) - E[\hat{f}_n])^2 + 2(f(x) - E[\hat{f}_n])(E[\hat{f}_n] - \hat{f}_n) + (E[\hat{f}_n] - \hat{f}_n)^2\right].$$

Taking the expectation,

$$E\left[(f(x) - E[\hat{f}_n])^2\right] + 2E\left[(f(x) - E[\hat{f}_n])(E[\hat{f}_n] - \hat{f}_n)\right] + E\left[(E[\hat{f}_n] - \hat{f}_n)^2\right]$$

From the above equation,

$$E[(E[\hat{f}_n] - \hat{f}_n)] = 0,$$

so it becomes

$$E\left[(f(x) - E[\hat{f}_n])^2\right] + E\left[(E[\hat{f}_n] - \hat{f}_n)^2\right]$$

$$E\left[(f(x) - E[\hat{f}_n])^2\right]$$

is the squared bias of the model,

$$\mathrm{Bias}^2 = (f(x) - E[\hat{f}_n])^2$$

the variance of the model

$$\mathrm{Var}(\hat{f}_n)$$

is represented by,

$$E\left[(E[\hat{f}_n] - \hat{f}_n)^2\right]$$

$$\mathrm{Var}(\hat{f}_n)$$

Therefore,

$$E\left[(f(x) - \hat{f}_n)^2\right] = \mathrm{Bias}^2 + \mathrm{Var}(\hat{f}_n)$$

Substituting it into the MSE equation,

$$\mathrm{MSE} = \mathrm{Bias}^2 + \mathrm{Var}(\hat{f}_n) + \tau^2$$