Name: CHAN King Yeung

SID: 1155119394

CSCI3230 Assignment 1

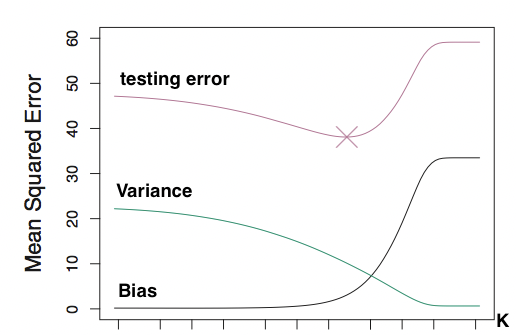
1. Learning rate is a tuning parameter in gradient descent which decides the step size at each iteration to search through the whole weight space. It ranges in between 0 and 1. Smaller requires more training epochs to make slow changes on weights at each iteration, whereas larger results in rapid changes with fewer training epochs. People often not to choose a large because a large may end up overshooting the minimum.

Chart

Description automatically generated

source: https://www.kaggle.com/residentmario/tuning-your-learning-rate

1. With a large number of layers in NN, the algorithm becomes more complex to calculate and update weights. Time cost would become a consider with an inordinately large number of layers. Moreover, a large number of layers may lead to overfitting. The NN has a huge information processing capacity to handle training data. When there is not enough training data for the NN, the NN will over learn (or process) from the training data and lead to overfitting. Besides, gradient descent may fail to update weights. After taking several partial derivatives, the gradient may vanish and will result to stop learning (or update) for weights.
2. In statistic, overfit a model would not affect the coefficient but also remain unbiased estimate. However, overfitting would lead to a model no longer adapt to a new environment. In other words, an overfit model may only favour to the training data. When there are unseen data, the predicted outcome may differ from the true outcome. More simply, it would form a lookup table instead of learning the pattern. Bias-variance trade-off is a possible solution.



source: Lecture from STAT4001

Given the fact that , we can control the variance to reduce the amount of bias. At the point where variance equals bias, the test error is minimised, such that the issues of overfitting can be avoided.