DSCI 508: Machine Learning

Mwamburi Bernard

HW02 Part 2

1. **Explain bias and variance. Explain the bias/variance trade off.**

Bias is a systematic error that occurs due to wrong assumptions in the machine learning process whereas **variance** is the amount by which the performance of a predictive model changes when it is trained on different subsets of the training data.

Bias/variance trade off occurs due to a tradeoff in model complexity - when a model's complexity is either too simple due to high bias and low variance(underfitting) or too complex due to high variance and low bias(overfitting).

Reference: *https://www.geeksforgeeks.org/bias-vs-variance-in-machine-learning/*

2. **What do you mean by the cost function?**

Cost Function is used to measure just how wrong the model is in finding a relation between the input and output based on the error between actual target values and predicted values and the determined coefficients and bias values. It tells you how badly your model is behaving/predicting.

Reference: *https://www.simplilearn.com/tutorials/machine-learning-tutorial/cost-function-in-machine-learning*

3. **For predicting continuous values, we use linear regression but for predicting classes we use logistic regression. Why do we use the sigmoid function for logistic regression?**

The sigmoid function is used in logistic regression to add nonlinearity to the data in the model. Since we are predicting classes, the relationship is nonlinear thus we must add nonlinearity to predict class.

Reference: *https://towardsdatascience.com/understanding-logistic-regression-the-odds-ratio-sigmoid-mle-et-al-740cebf349a3*

4. **What is the difference between Maximum likelihood and negative log likelihood? Why do we take negative log of the cost function in logistic regression?**

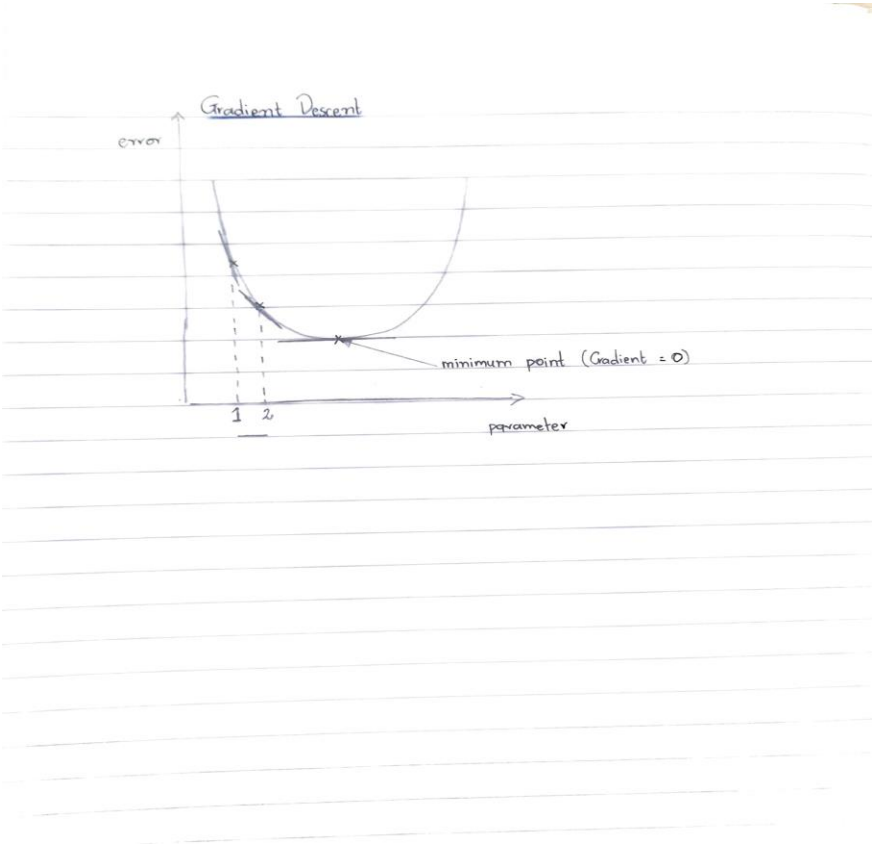
Likelihood is the chance of an event occurring in a set of different events whether dependent or independent in a probability distribution, the **maximum likelihood** is then obtained by maximizing the likelihood function. The **negative log likelihood is** then the negative of the natural logarithm of the likelihood which is then used as the loss function for a logistic regression model.

We use the negative log of the cost function as computers are not capable of exact numeric representation. You may lose precision in multiplying numbers together if the numbers are too high or too low.

Reference: *https://medium.com/deeplearningmadeeasy/negative-log-likelihood-6bd79b55d8b6*

5. **Gradient descent is an optimization algorithm through which the hyperparameters of your hypothesis are updated, based on the cost function. Can you explain how parameters are updated with gradient descent?**

Parameters are iteratively adjusted in the direction of negative gradient of the cost function until the minimum is reached. The change in the gradient is normally controlled by the **learning rate.**



Reference: *https://www.geeksforgeeks.org/gradient-descent-algorithm-and-its-variants/*

6. **What are the benefits of using k-fold cross validation?**

- It helps prevent overfitting by providing a more robust estimate of the model's performance on unseen data.

- It can be used to compare different models and select the one that performs the best.
- K-fold cross validation can be used to optimize hyperparameters of a model by selecting values that result in the best performance.
- By using all the available data for both training and validation, cross validation is more data-efficient compared to traditional validation techniques.

Reference: *https://www.geeksforgeeks.org/cross-validation-machine-learning/?ref=header_search*