

Macalla, Anel Thom T.
Assignment #4

1. Explain the Bias-Variance Trade-off

The bias–variance tradeoff is the property of a model that the variance of the parameter estimates across samples can be reduced by increasing the bias in the estimated parameters. The bias–variance dilemma or bias–variance problem is the conflict in trying to simultaneously minimize these two sources of error that prevent supervised learning algorithms from generalizing beyond their training set:

- The bias error is an error from erroneous assumptions in the learning algorithm. High bias can cause an algorithm to miss the relevant relations between features and target outputs (underfitting).
- The variance is an error from sensitivity to small fluctuations in the training set. High variance can cause an algorithm to model the random noise in the training data, rather than the intended outputs (overfitting).

The bias–variance decomposition is a way of analyzing a learning algorithm's expected generalization error with respect to a particular problem as a sum of three terms, the bias, variance, and a quantity called the *irreducible error*, resulting from noise in the problem itself.

2. Why Model Optimization is Important?

Because optimized solutions represent the best compromise between a number of inter-related variables to achieve a specific goal, it's often impossible to determine the right solution by any other means.

3. What is Cross-Validation? Is it applicable only to classification algorithms? Why is it important?

- *What is Cross Validation?*

Cross-validation is a statistical method used to estimate the skill of machine learning models. It is commonly used in applied machine learning to compare and select a model for a given predictive modeling problem because it is easy to understand, easy to implement, and results in skill estimates that generally have a lower bias than other methods.

- *Is it applicable only to classification algorithms?*

No

- *Why is it important?*

It is important because Cross-Validation is an essential tool in the Data Scientist toolbox. It allows us to utilize our data better. Before I present you my five reasons to use cross-validation, I want to briefly go over what cross-validation is and show some common strategies.

4. What is Ensemble Method?

Ensemble methods are techniques that create multiple models and then combine them to produce improved results. Ensemble methods usually produces more accurate solutions than a single model would. This has been the case in a number of machine learning competitions, where the winning solutions used ensemble methods.

References:

https://en.wikipedia.org/wiki/Bias%E2%80%93variance_tradeoff

<https://www.riverlogic.com/blog/optimization-modeling-everything-you-need-to-know#:~:text=The%20Value%20of%20Optimization%20Modeling&text=Because%20optimized%20solutions%20represent%20the,solution%20by%20any%20other%20means.>

<https://towardsdatascience.com/cross-validation-430d9a5fee22>

<https://www.toptal.com/machine-learning/ensemble-methods-machine-learning>