# Questions and Answers for Bias-Variance Decomposition, Ensemble Method I

### Bingzhang Zhu

### October 27, 2021

## 1 What is Bias–variance tradeoff?

In statistics and machine learning, the bias–variance tradeoff is the property of a model that the variance of the parameter estimated 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 may result from an algorithm modeling the random noise in the training data (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.

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

## 2 What does random refer to in 'Random Forest'?

**Random Record Selection**: Each tree in the forest is trained on roughly 2/3rd of the total training data (exactly 63.2%) and here the data points are drawn at random with replacement from the original training dataset. This sample will act as the training set for growing the tree.

**Random Variable Selection**: Some independent variables(predictors) say, m are selected at random out of all the predictor variables, and the best split on this m is used to split the node.