**Bagging on Low Variance Models.**

Bagging (also known as bootstrap aggregation) is a technique in which we take multiple samples repeatedly with replacement according to uniform probability distribution and fit a model on it. It combines multiple predictions to give a better prediction by majority vote or taking the aggregate of the predictions. It is a known fact that this technique is highly effective on models which tend to overfit on the dataset (high variance models) like.

If this technique is so good, why do we use it only on models which show high variance? What happens when we use it with models which have low variance? Let us try to understand the underlying issue with the help of a demonstration.

I’ll be using bagging on decision tree to prove that bagging improves the accuracy for high variance models and compare it to bagging on linear regression which is a stable model when there is a linear trend between the predictor and the target variables.

**Bias and Variance**

We will be talking about Bias and Variance throughout the article so let us get an idea of what it is first.

High bias refers to the oversimplification of the model. i.e. When we are unable to capture the true relation of the training dataset. Our objective of creating a model is to capture the true nature of the dataset which makes high bias an undesired phenomenon.

<Image of High bias>

High variance refers to the situation when we are overcomplicating our model. i.e. The situation where, in the process of capturing the true nature of the model, we are creating a model which learns the training data so well that its accuracy deteriorates on any other dataset. This situation is also undesired as our objective is to make predictions for unseen data.

<Image of high variance>

When we are creating a model, we want to strike a balance between the bias and variance. Bias and Variance are opposite of each other so whenever we try to reduce the variance, we are increasing the bias of the model at the same time. This dilemma of overfitting/underfitting is called Bias-Variance Tradeoff. This image gives a good idea of their relation witch each other.

<Image of bias-variance tradeoff>

**High Variance Model – Decision Tree**

Demonstrate how it works well with bagging.

Talk about linear regression

Demonstrate that it does not work well

Why does linear regression doesn’t work well with bagging. And what happens

Conclude with intuition for bagging.

References:

* Class notes
* <https://www.analyticsvidhya.com/blog/2018/06/comprehensive-guide-for-ensemble-models/>
* Data Mining and Predictive Analytics – Daniel T. Larose, Chantal D. Larose