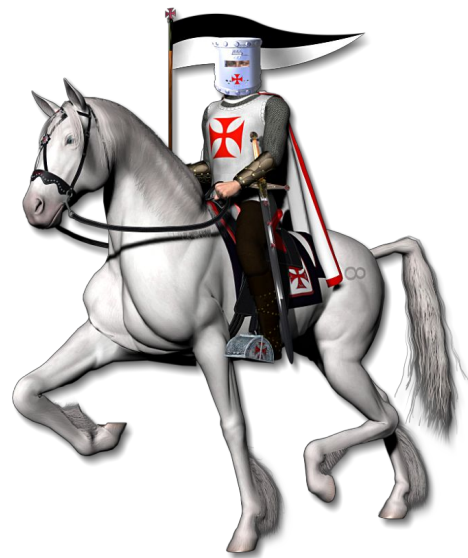


Bias - Variance Tradeoff

Why Bias and Variance are never
close together?



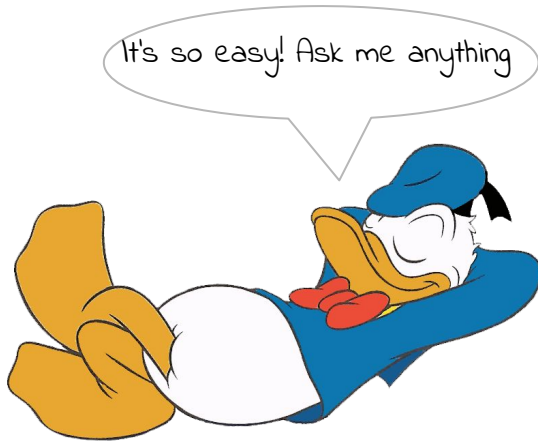
What is Bias?

- The amount by which the expected model prediction differs from the true value of target or how far off our predictions are from real values



Bias

- Model with high bias pays very little attention to the training data and oversimplifies the model



Bias

- It always leads to high error on training and test data

You have \$1000 bank balance. Go shopping and treat yourself!



Oops! My bad
~_(\ツ)_/~

What is Variance?

- The amount by which the model prediction would change if we estimated it using a different training dataset

The results are out! You failed



I've tried different method. You scored full

Girl power, Yeah!



Variance

- Model with high variance pays a lot of attention to training data and doesn't generalize on the data which it hasn't seen before



Variance

- As a result, such models perform very well on training data but has high error rates on test data

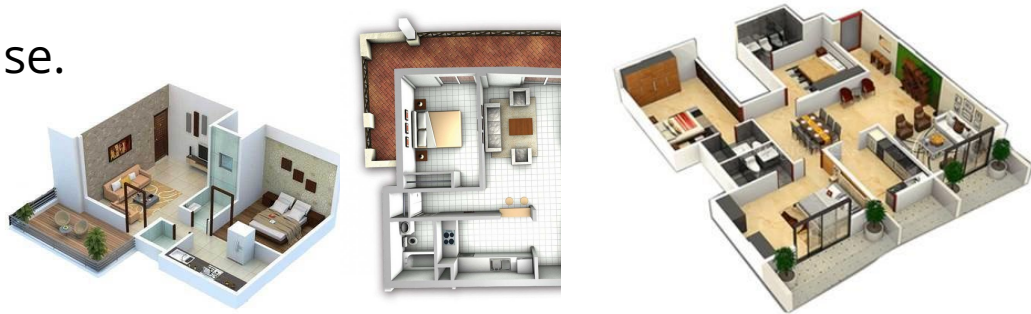


Exam



Noise

- If you were to model the price of an apartment, you know that the price depends on the area of the apartment, no. of bedrooms, etc.
- So those factors contribute to the pattern — more bedrooms would typically lead to higher prices.
- However, all apartments with the same area and no. of bedrooms do not have the exact same price.
- The variation in price is the noise.



Noise

- Consider driving.
- Given a curve with a specific curvature, there is an optimal direction of steering and an optimal speed.
- When you observe 100 drivers on that curve, most of them would be close to that optimal steering angle and speed.
- But they will not have the exact same steering angle and speed.
- So again, the curvature of the road contributes to the pattern for steering angle and speed, and then there is noise causing deviations from this optimal value.



Archery Competition



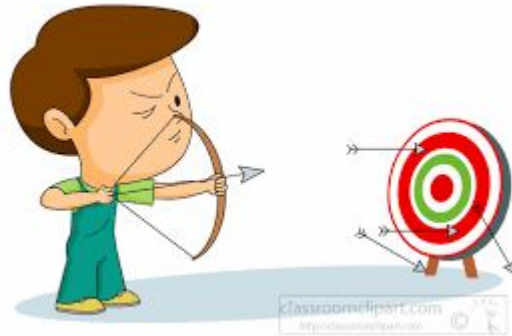
Perfect Model

- Bull's eye!



Other possibilities

- Try hitting away from centre:
- That's how much your prediction(model) got worse.



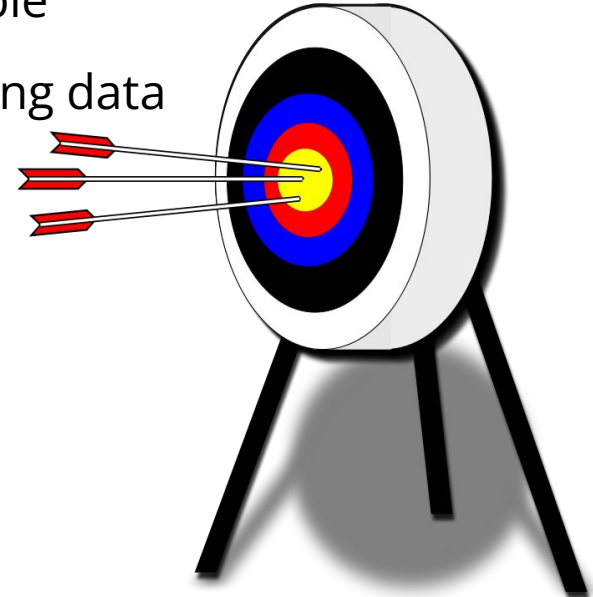
Other Possibilities

- Imagine we can repeat our entire model building process to get a number of separate hits on the target.
- Each hit represents an individual realization of the model



Low Bias & Low Variance

- This would have to be the ideal situation which we would like for our Machine Learning model
- The error of prediction should be as less as possible
- Predictions shouldn't change with change in training data



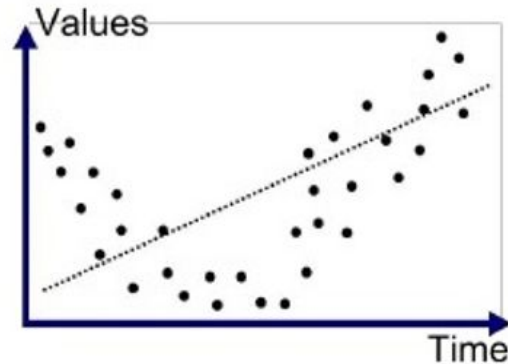
High Bias & High Variance

- This would have to be the worst situation which we would like for our Machine Learning models
- The error of prediction or bias is very high
- The predictions also fluctuate rapidly with change in training data



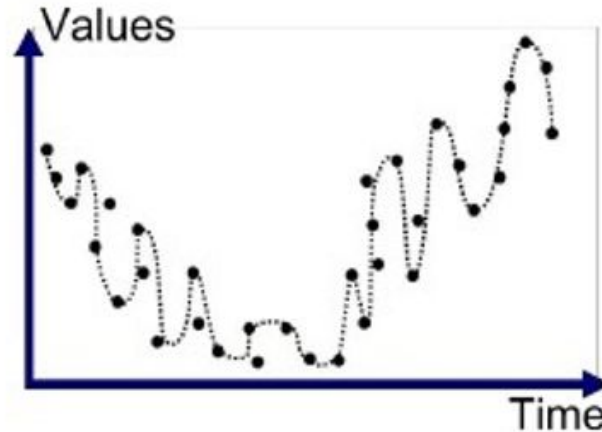
High Bias & Low Variance

- Underfitting happens when model is unable to capture the underlying pattern of data
- Happens when we have very less amount of data to build an accurate model
- Also happens when we try to build a linear model with non-linear data
- Eg.: Linear, Logistic Regression



Low Bias & High Variance

- Overfitting happens when our model captures the noise around the underlying pattern in data
- Happens when we train our model a lot on a noisy dataset
- Happens in complex models like decision trees, SVMs



Overfitting

- The goal of machine learning is to model the pattern and ignore the noise.
- Anytime an algorithm is trying to fit the noise in addition to the pattern, it is overfitting.

Why Bias-Variance Tradeoff

- If our model is too simple and has very few parameters, then it may have high bias and low variance.
- On the other hand if our model has large number of parameters then it's going to have high variance and low bias
- So we need to find the right/good balance without overfitting and underfitting the data.

Managing Bias & Variance

A proper Machine Learning workflow includes:

- Separate training and test data sets
- Trying appropriate algorithms
- Fitting model parameters
- Tuning impactful hyperparameters
- Proper performance metrics
- Systematic cross-validation



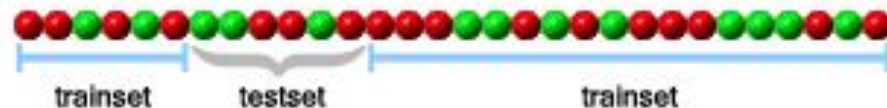
Be a good boy

ONE ITERATION OF A 5-FOLD CROSS-VALIDATION:

1-ST FOLD:



2-ND FOLD:



3-RD FOLD:



4-TH FOLD:



5-TH FOLD:

