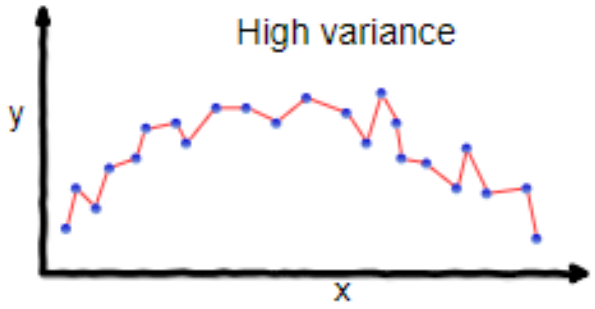
Logistic regression

Bias vs Variance

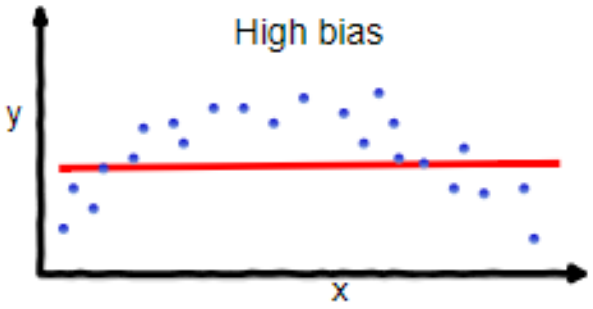
- Whenever we discuss model prediction, it’s important to understand prediction errors:
  - bias and
    - Difference between the average prediction of our model and the correct value which we are trying to predict.
  - Variance
    - Variance is the algorithm’s tendency to learn random things irrespective of the real signal.

What is the TRADE-OFF?

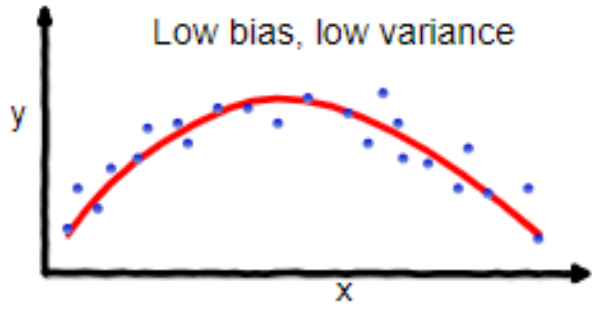
- If our model is too simple and has very few parameters then it may have high bias and low variance.
- On the other hand if our model has large number of parameters then it's going to have high variance and low bias.
- So we need to find the right/good balance without overfitting and underfitting the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



overfitting



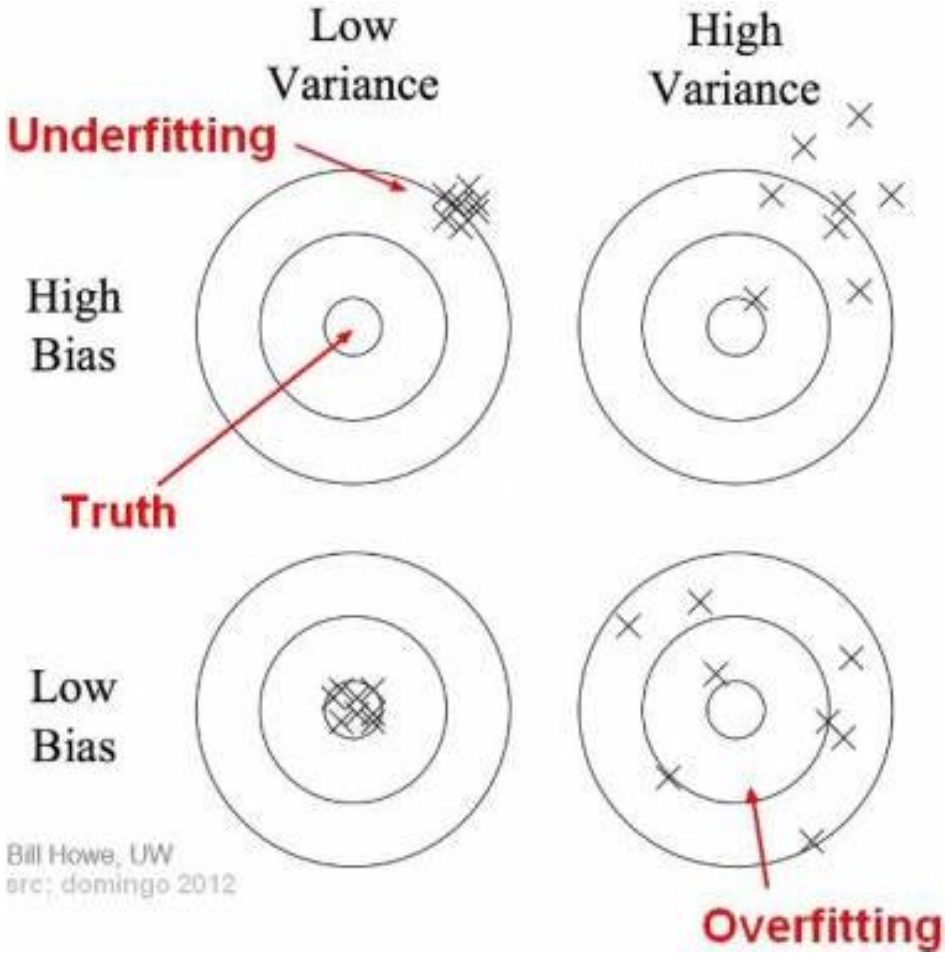
underfitting



Good balance

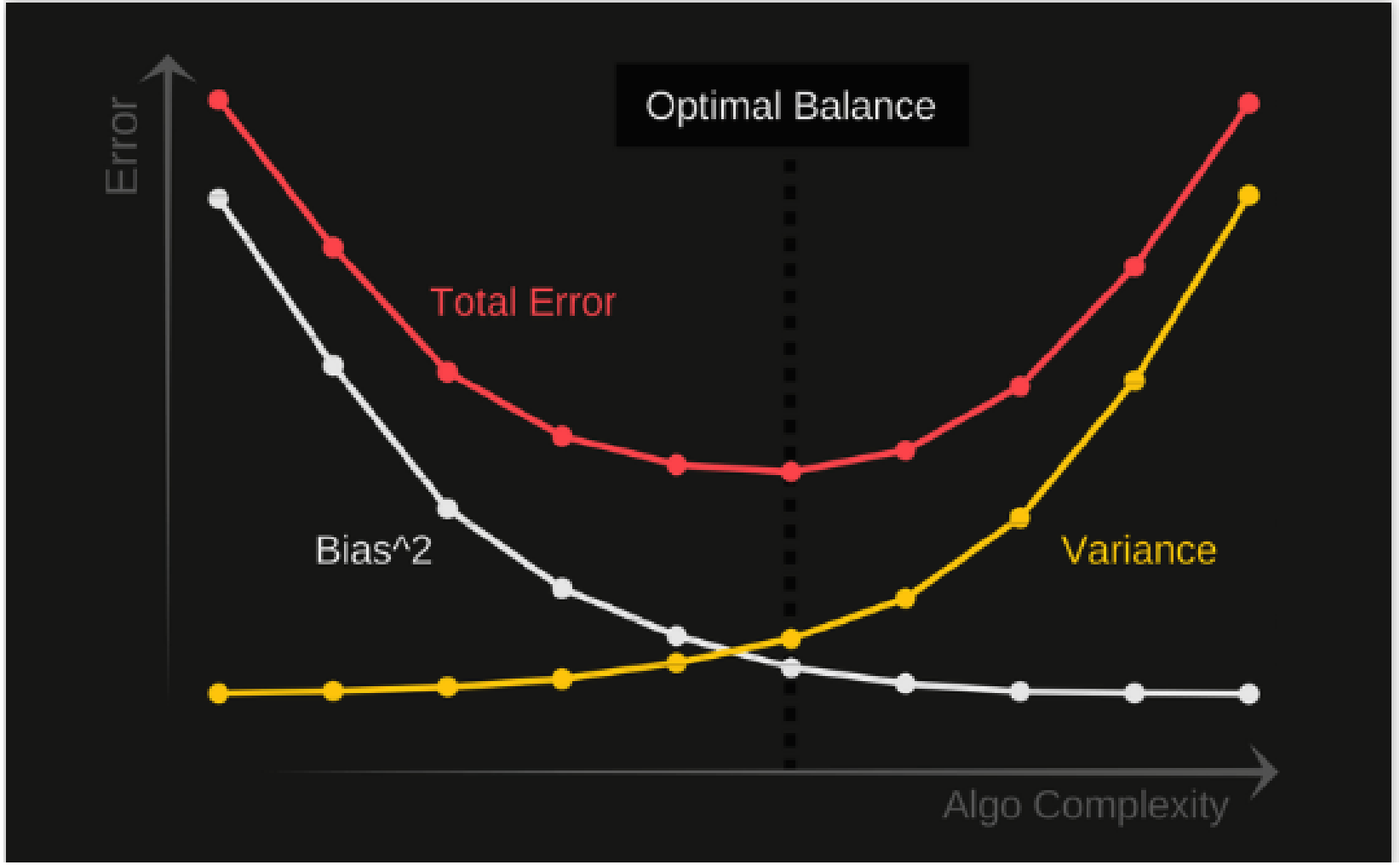
Bias and variance using bulls-eye diagram

- In the diagram, center of the target is a model that perfectly predicts correct values.
- As we move away from the bulls-eye our predictions become get worse and worse.
- Underfitting happens when a model unable to capture the underlying pattern of the data.
- Overfitting happens when our model captures the noise along with the underlying pattern in data. It happens when we train our model a lot over noisy dataset.



Total Error

- To build a good model, we need to find a good balance between bias and variance such that it minimizes the total error.



If you have HIGH VARIANCE PROBLEM

- You can get more training examples because a larger the dataset is more probable to get a higher predictions.
- Try smaller sets of features (because you are overfitting)

If you have HIGH BIAS PROBLEM

- Try getting additional features, you are generalizing the datasets.
- Try adding polynomial features, make the model more complicated.

Thanks