Bias-Variance Trade off

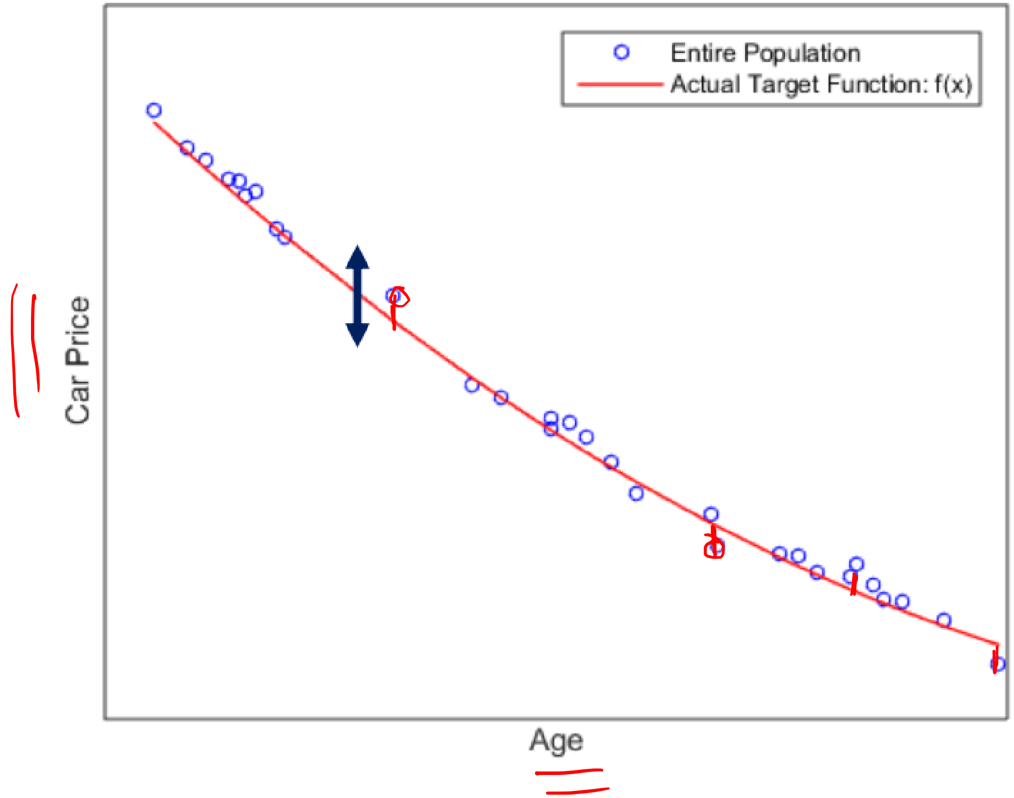
Dr. Muhammad Wasim

Noise

- Noise is the irreducible error inherent in the data.

$$y = f(x) + \epsilon$$

- This noise is the property of the data and has nothing to do with the model.
- E.g. the relationship between a car's price and its age is not a perfect relationship.
- No model, can capture the exact relationship.



Why discuss Bias and Variance?

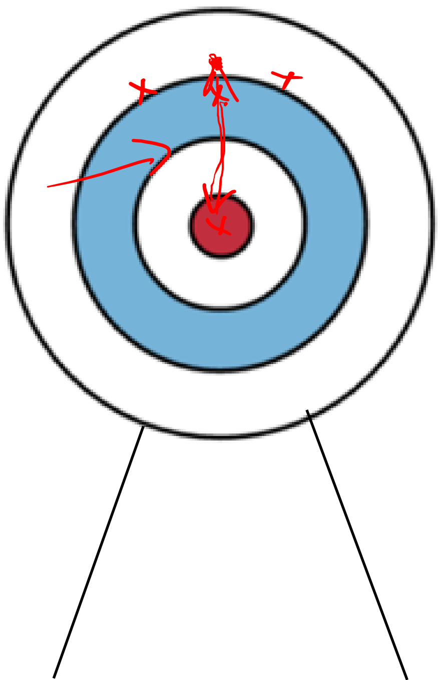
- Bias-Variance Decomposition is a key component in understanding learning algorithms.
- Understanding how different sources of error lead to bias and variance helps us improve the data fitting process resulting in more accurate models.
- Helps understand and avoid overfitting and underfitting.
- Helps explain why simple models can outperform the more complex ones.
  - A regression model with fewer parameters maybe better than one with more parameters.
  - A neural network model with fewer neurons maybe better than one with more neurons.
  - A simple classifier such as Naïve Bayes maybe better than decision trees.

Bais-Variance Trade-off

- We assume we could repeat the whole model building process more than once: each time we gather new data and run a new analysis creating a new model.
- Due to randomness in the underlying data sets, the resulting models will have a range of predictions.
- You will have different predictions for your target for the different models.

- Bias measures how far off *in general* these models' predictions are from the correct value.
- The variance is how much the predictions for a given point vary between different realizations of the model.
- Note that variance has nothing to do with where the actual target is.

bias



3 datasets

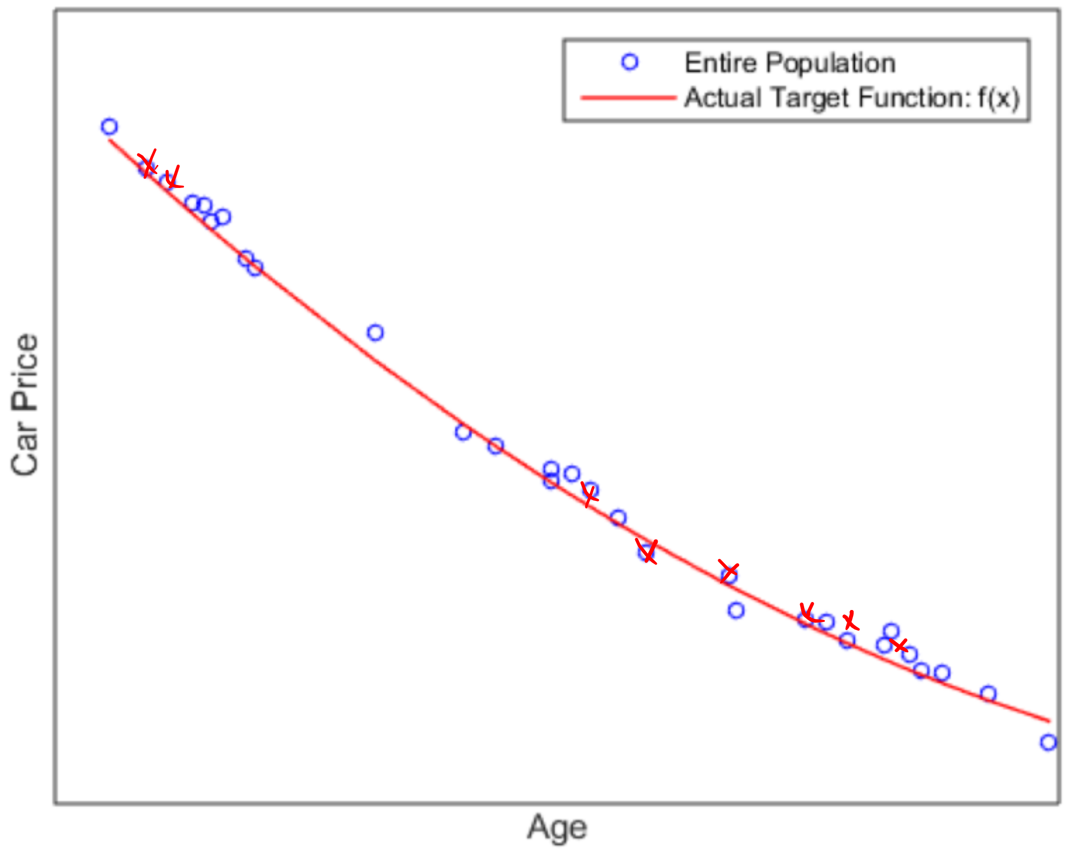
1 model A
2 model A
3 model A

New Example

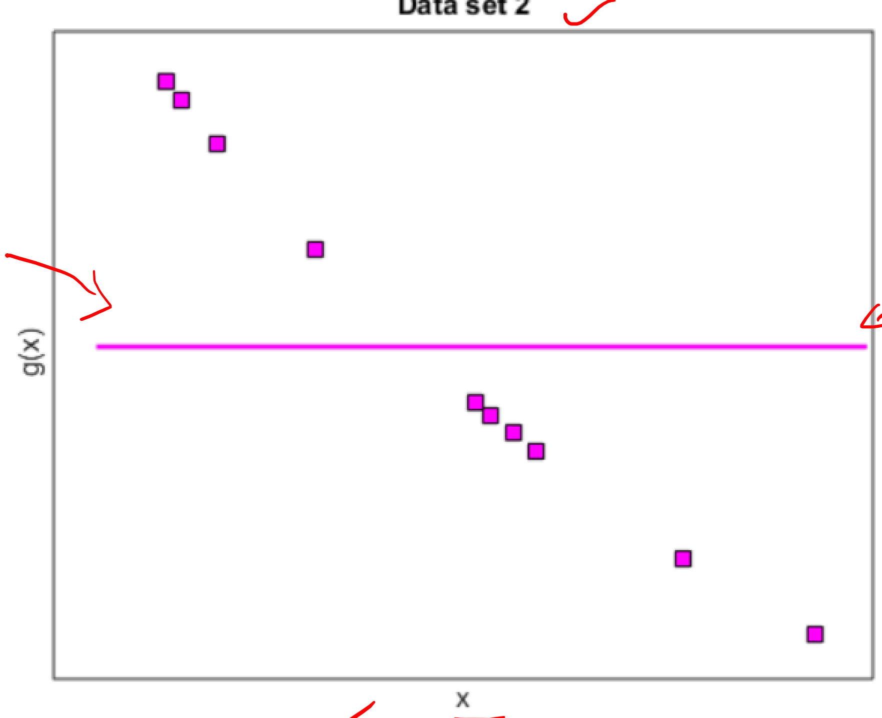
Bias-Variance Trade off - II

Example

- The objective is to create a model that predicts the price of a car based on its Age.
- The red curve denotes the underlying relationship between the Age and the price of cars in the entire population.

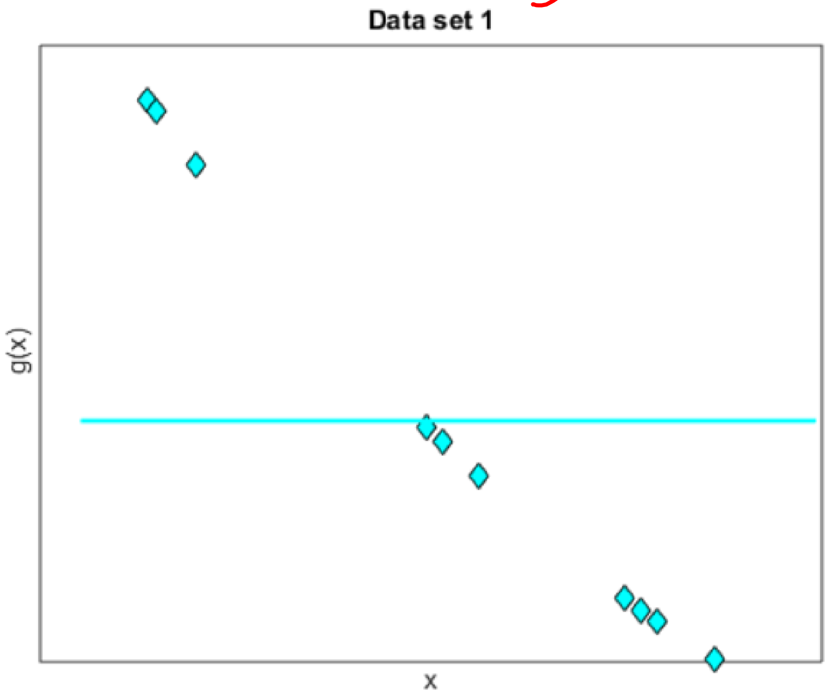


- Assume different groups collect different samples and create a simple constant model based on their data set.
- Every data set results in a slightly different line g(x).
- The predicted hypotheses g(x) for a data set whose cars are worth below the true relationship, is different from a data set where most cars are worth more than the typical values in the population.

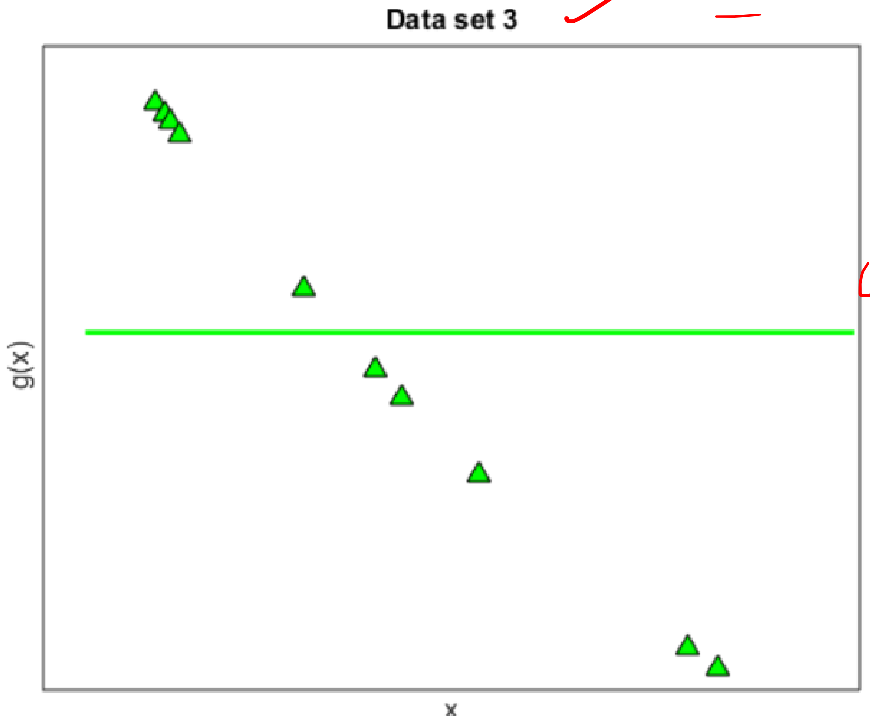


h_1

$y = 4$

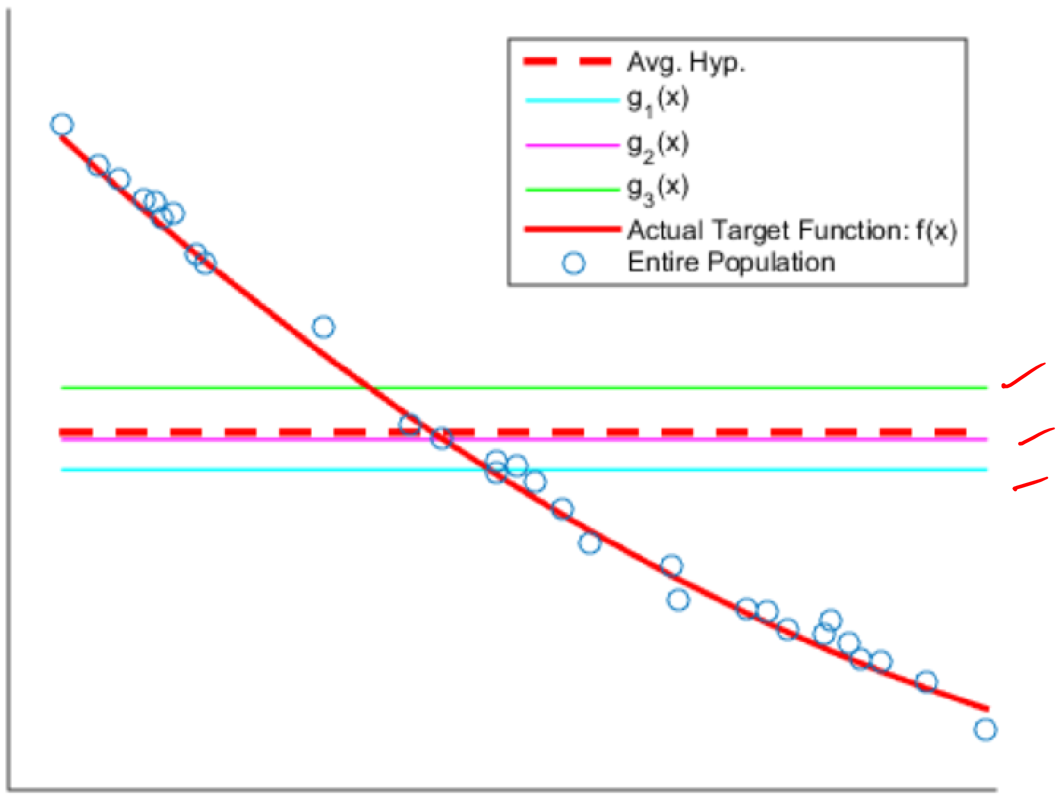


h_3

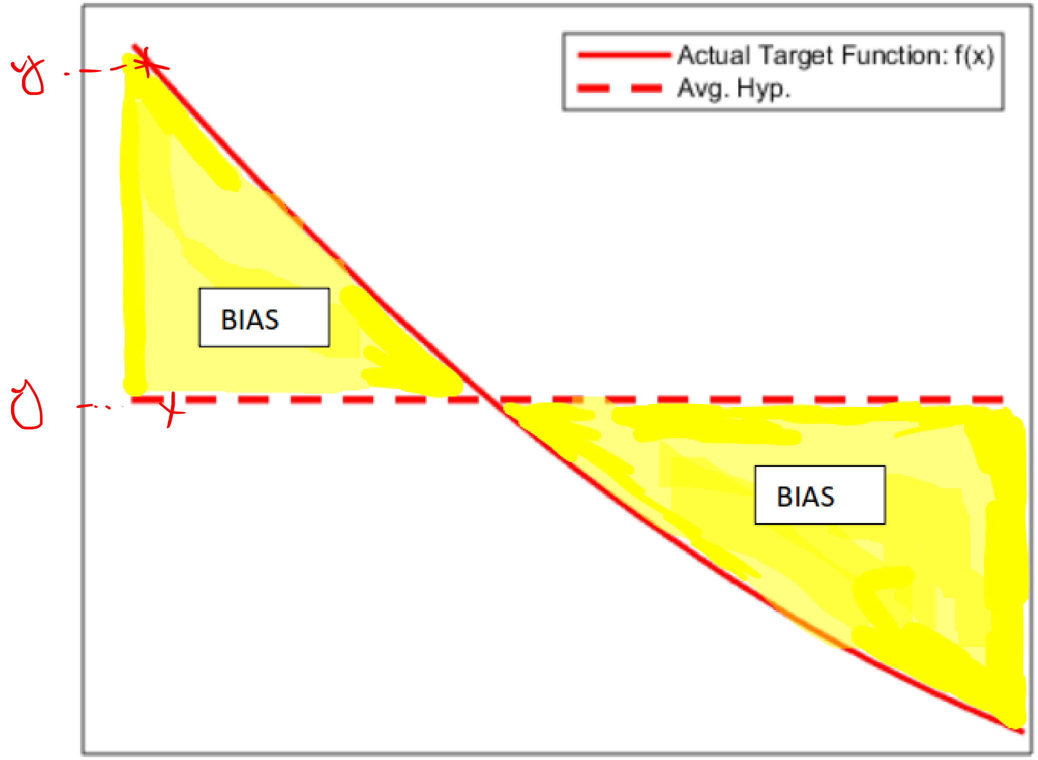


h_2

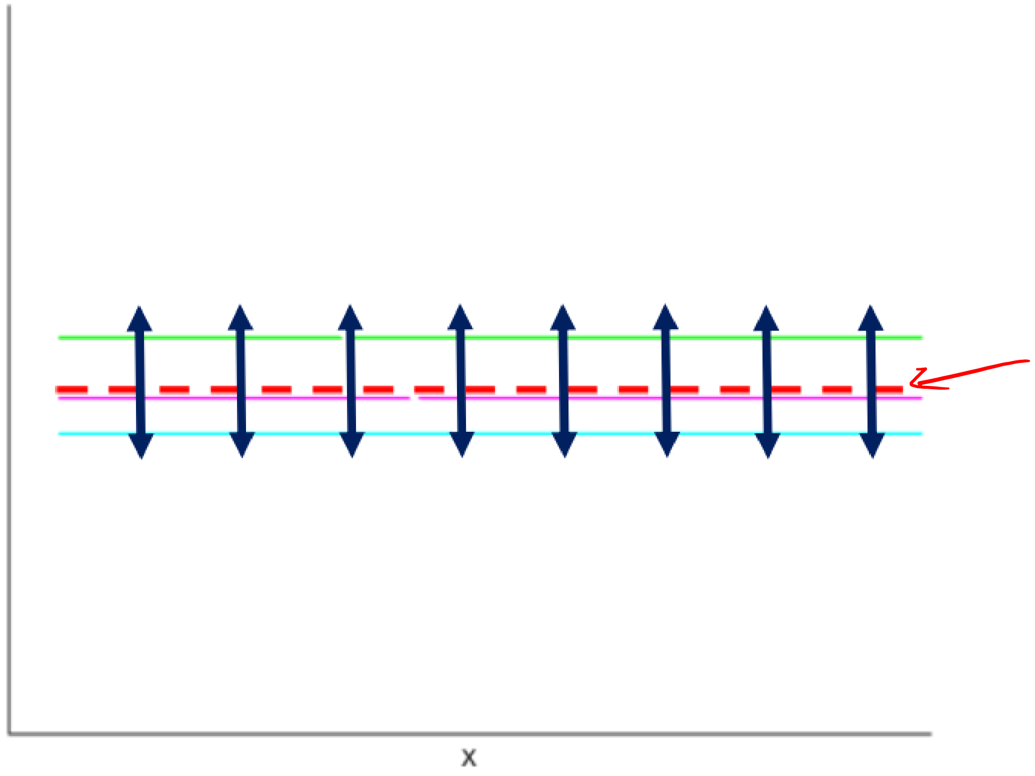
Average Hypothesis - Conceptually



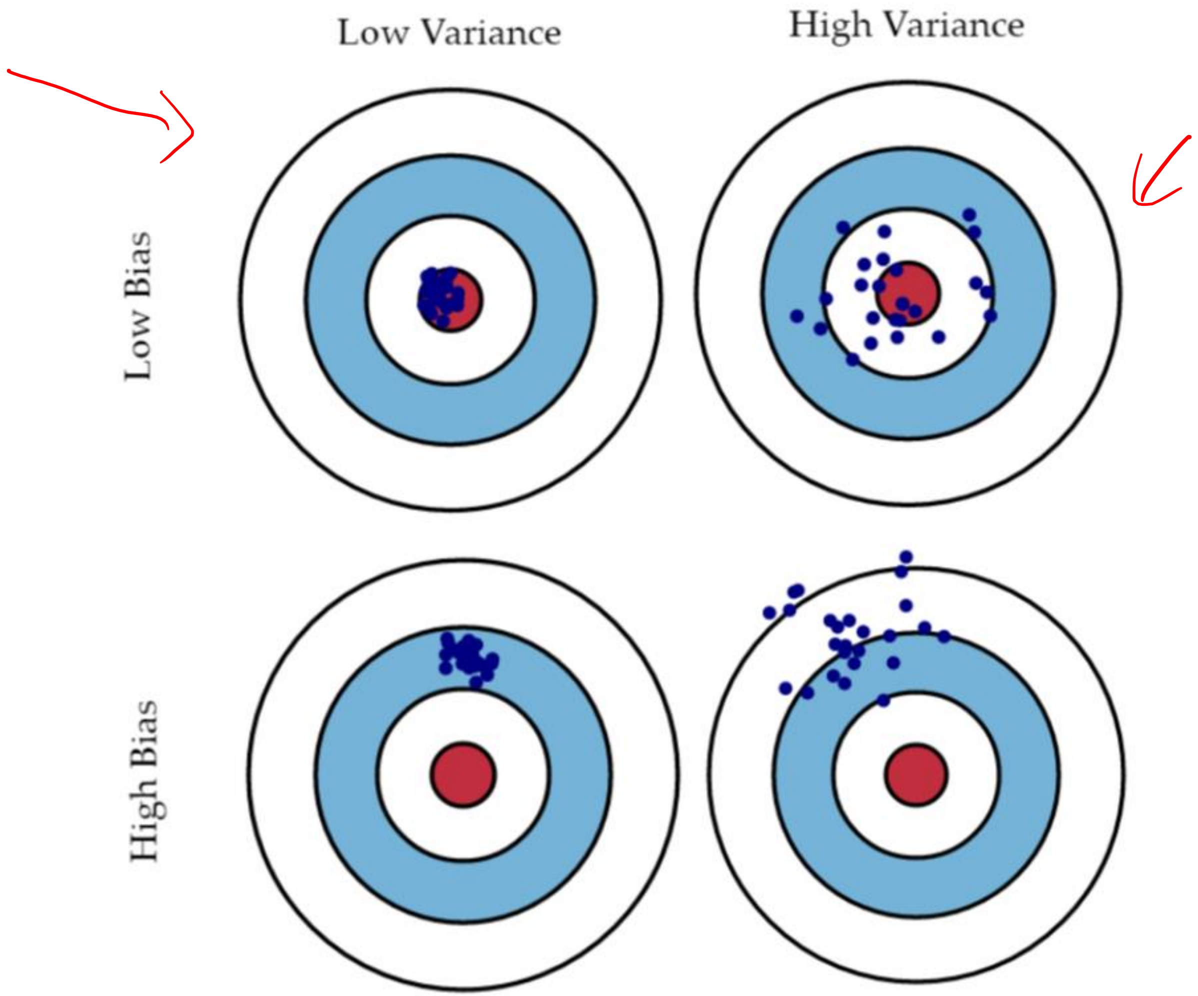
Bias - Conceptually



Variance - Conceptually



Graphical Illustration of Bias and Variance

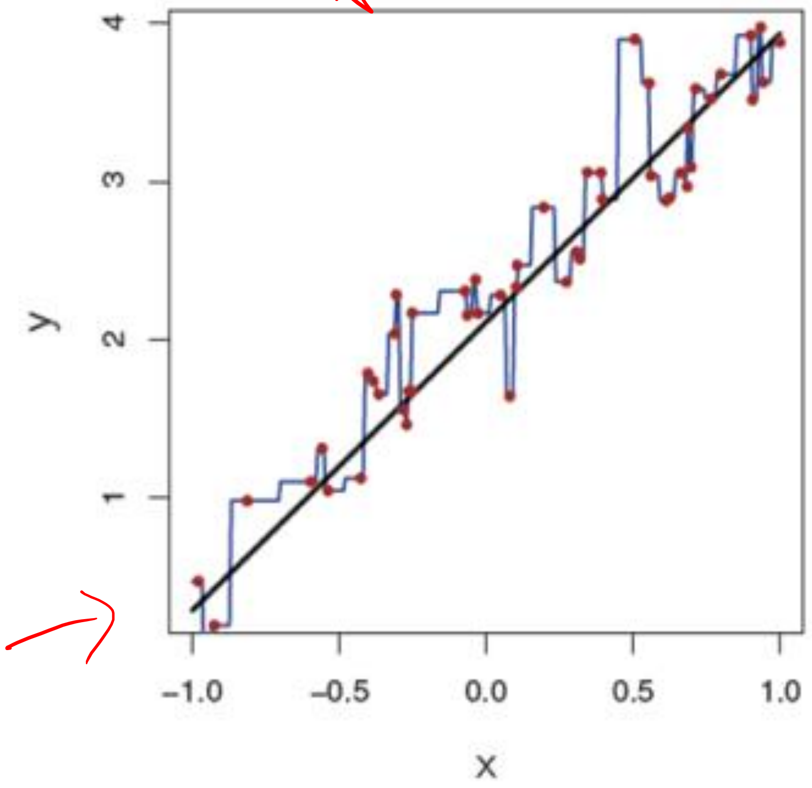


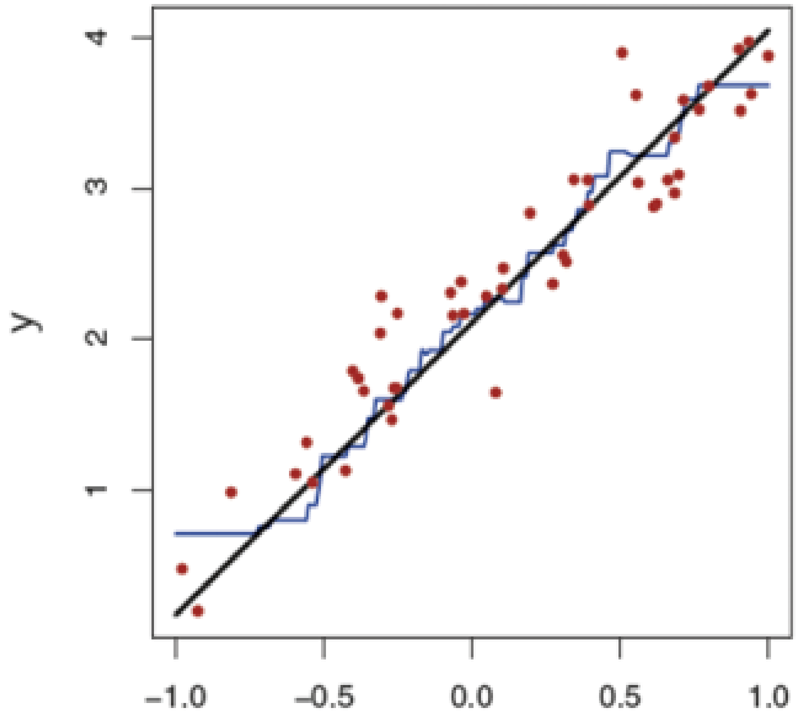
Bias-Variance Trade off - III

Example 1 – KNN Algorithm

↓Variance
↓Bias

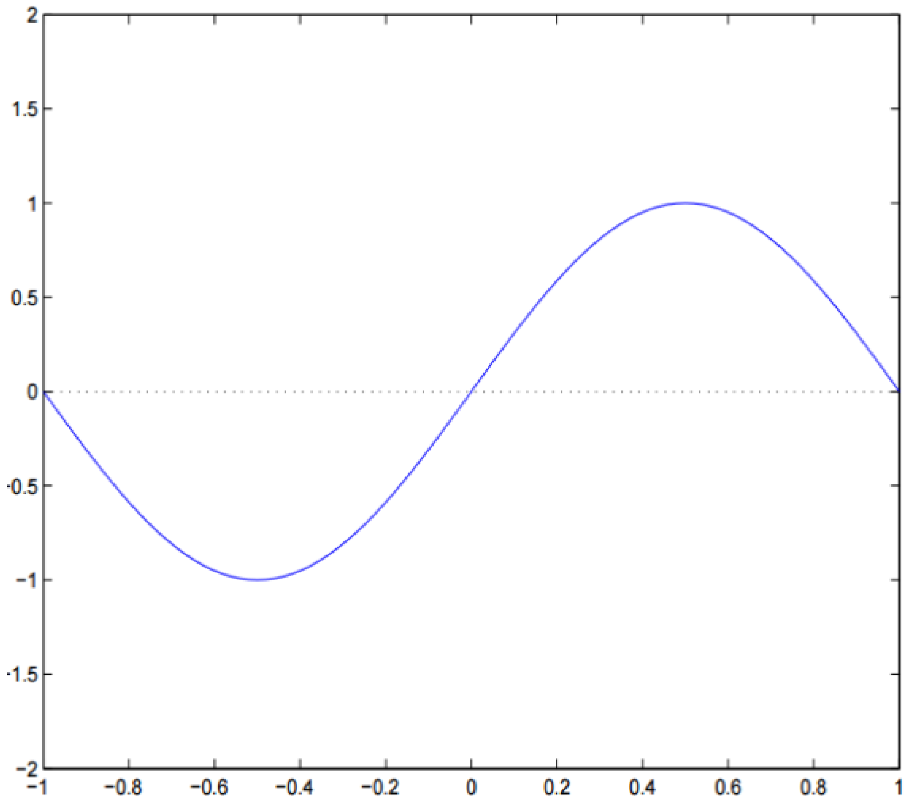
↑Variance
↑Bias





The model with one vs nine neighbors

Example 2 – Approximating a sinusoid function

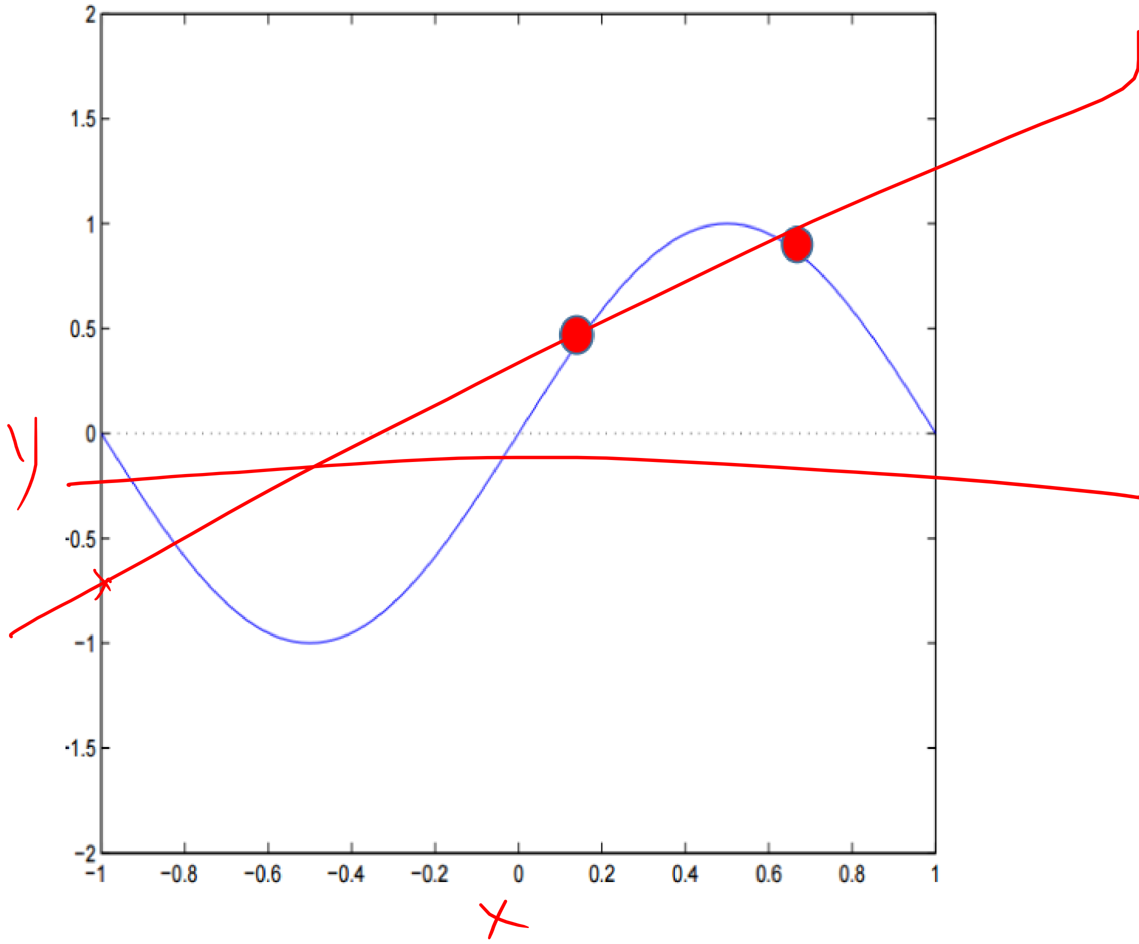


Constant model vs. Linear model

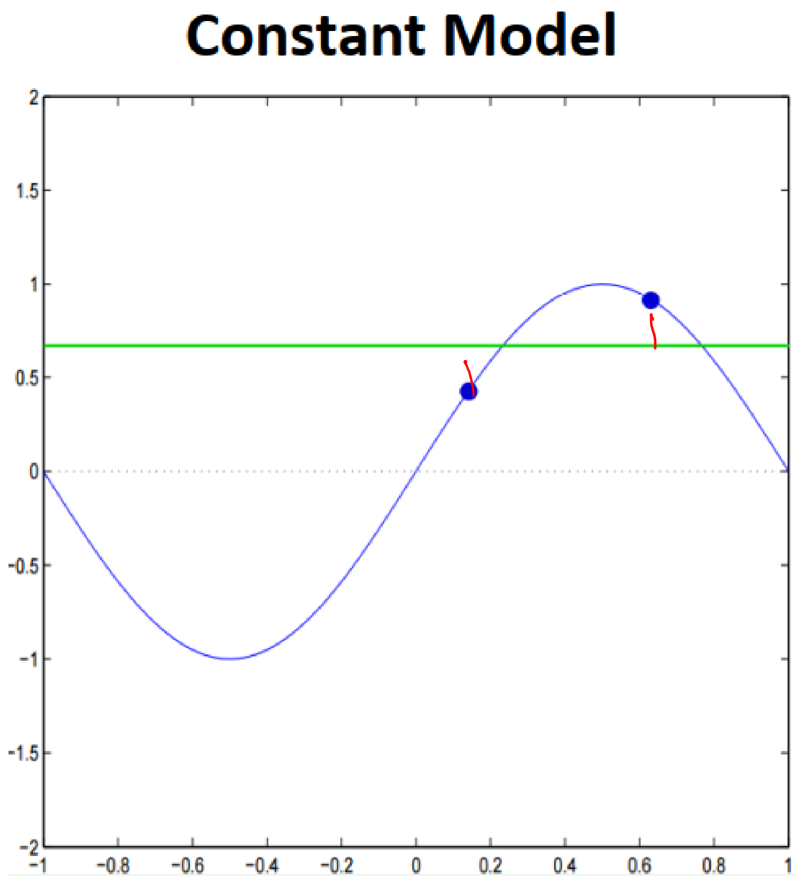
Example 2 – Approximating a sinusoid function (cont.)

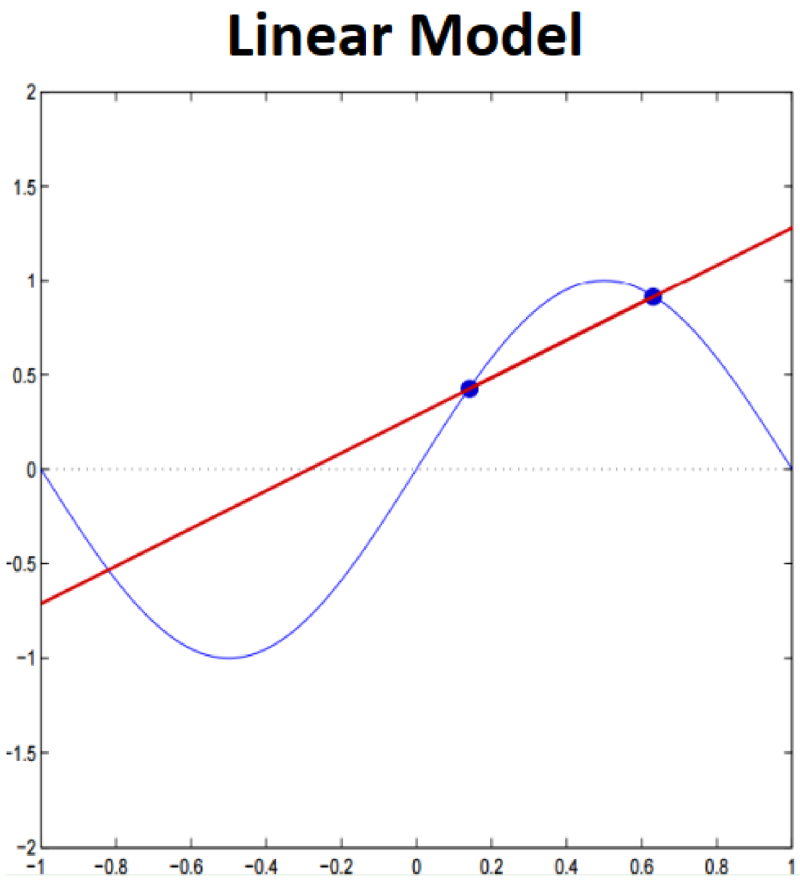
- You **don't know** the target function.
- You must use your data set of size N=2 to learn the target function.
- Your hypotheses sets are constant and linear models, i.e.

1. $h(x) = b$ ✓
2. $h(x) = mx + b$



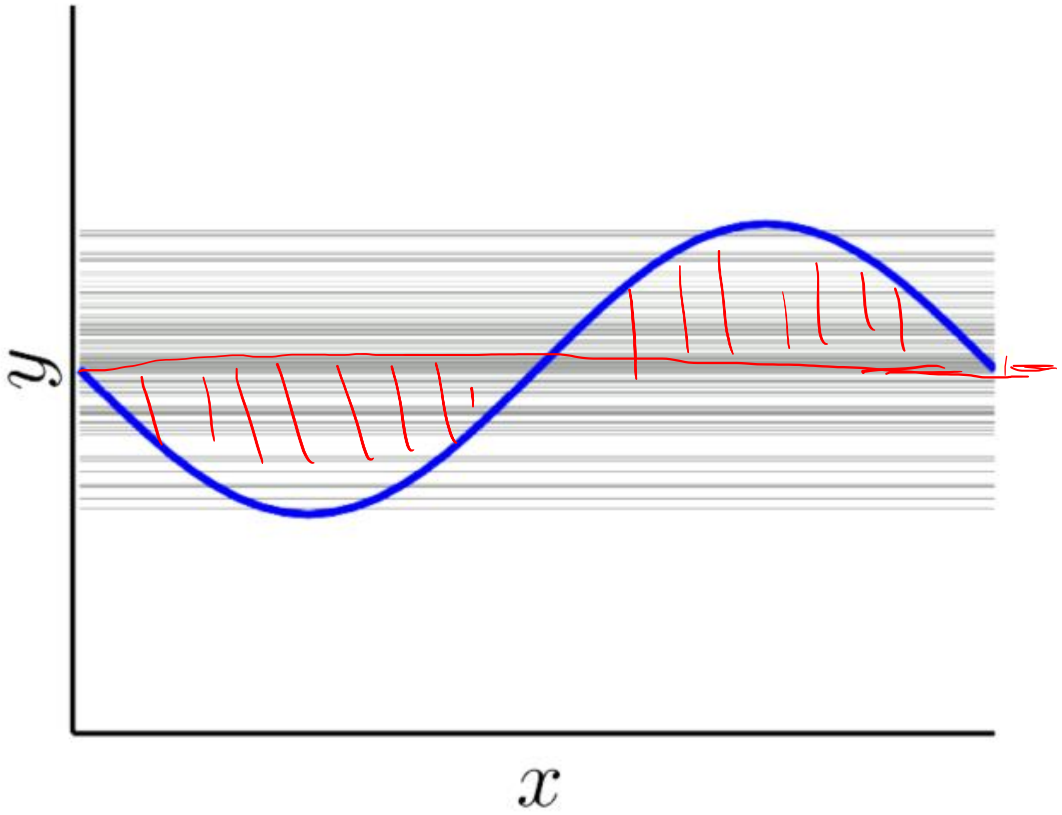
Example – cont.





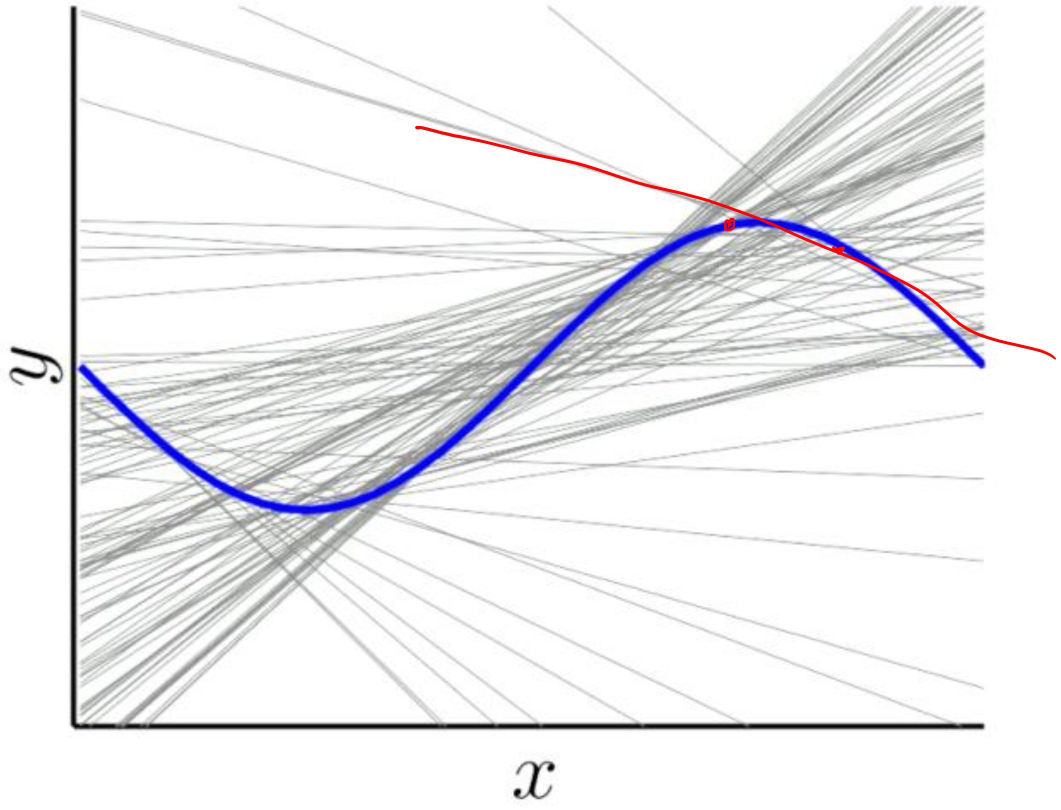
Repeating the Model Building with Different Data Sets

$y = b$



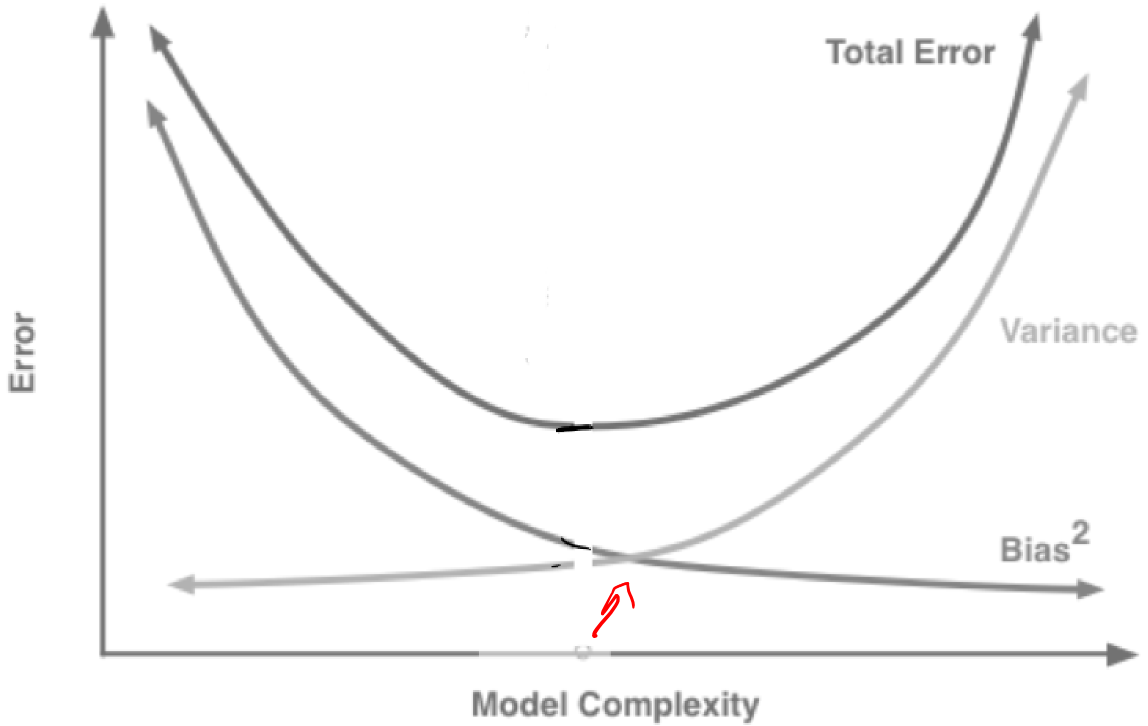
↑ Bias ⟷ ↓ Variance

$y = b + mx$



↓ Bias ⟷ ↑ Variance

Bias – Variance Trade Off Plot



Can we assess if our model has high bias or variance with experiments?

Train set accuracy: 99%

Test set accuracy: 90%

High variance / Overfitting – It means you need to increase dataset or decrease model complexity.

Train set accuracy: 85%

Test set accuracy: 84%

High Bias / Underfitting – It means you need to use a more complex model so relationship can be modeled properly.