Model Evaluation

DR. BHARGAVI
PROFESSOR
VIT CHENNAI

Introduction

- Suppose that you train a ML Model on a dataset and you get a very low training error.
- Now, the same trained model is used to predict the test data and it gives very high error.
- Low training error does not imply low test error.
- What should we do?
  - Increase training examples
  - Reduce/Increase the features
  - Add more polynomial features (In the case of linear regression)
  - Increase/decrease regularization parameter

Measuring the Quality of Fit

- Regression – Mean Squared Error (MSE)

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{f}(x_i))^2$$

- Classification - Misclassification error (aka 0/1 misclassification error)

$$Error = \frac{1}{n}\sum_{i=1}^{n} err(h_x(\boldsymbol{x}_i), y_i)$$

Where err(h_w(x),y) = 1 if h_w(x) >= 0.5 and y = 0 or h_w(x) < 0.5 and y = 1 (misclassified)

0 Otherwise (Correct classification)

Model Assessment And Model Selection

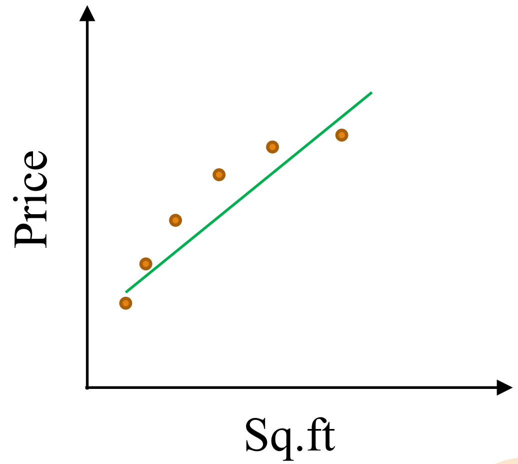
Model Assessment

The process of evaluating a model's performance is known as model assessment

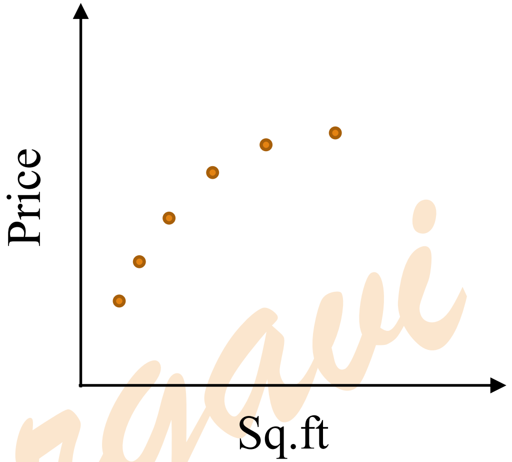
Model Selection

The process of selecting the proper level of flexibility for a model is known as model selection (Ex: Degree of the polynomial in linear regression)

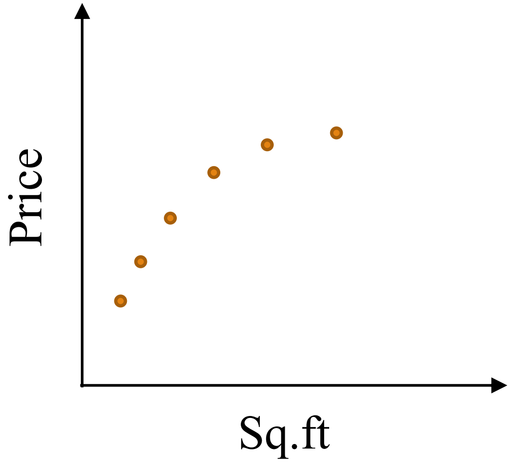
Model Selection - Train/Validation/Test Sets



$$h(x) = w_0 + w_1 x$$



$$h(x) = w_0 + w_1 x + w_2 x^2$$



$$h(x) = w_0 + w_1 x + \ldots + w_{10} x^{10}$$

Model Selection - Train/Validation/Test Sets

- Split the dataset into (in the proportion like 60:20:20 or 80:10:10)
  - Training data
  - Validation data
  - Test data
- Use Training data to optimize the model parameters i,e w's
- Use Validation data to find the degree of polynomial with least error
- Estimate the generalization error using the test set.

Three sources of error

Expected Test MSE, for a given test data x_0 can be written as

$$E\left(y_0 - \hat{f}(x_0)\right)^2 = \mathrm{Var}(\hat{f}(x_0)) + [\mathrm{Bias}(\hat{f}(x_0))]^2 + \mathrm{Var}(\epsilon)$$

Var(ϵ) is the **irreducible error** (noise).

Bias: Error that is introduced by approximating a real-life problem, which may be extremely complicated, by a much simpler model.

Variance: Variance refers to the amount by which $\hat{f}$ would change if we estimated it using a different training data set.

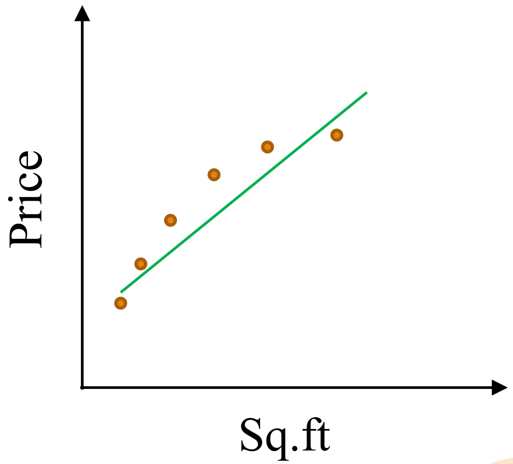
In order to ***minimize the expected test error***, we need the ML method that can simultaneously achieves ***low variance and low bias***.

Three sources of error

Variance

- Since the training data are used to fit the machine learning method, different training data sets will result in a different $\hat{f}$.
- But ideally the estimate for f should not vary too much between training sets.
- However, if a method has high variance then small changes in the training data can result in large changes in $\hat{f}$.
- In general, more flexible Machine Learning methods have higher variance.
- In a more flexible model, changing any one of the data points may cause the estimate $\hat{f}$ to change considerably (since the model tries to fit the data correctly).
- In a less flexible model, changing any one of the data points may not cause the change to estimate $\hat{f}$.

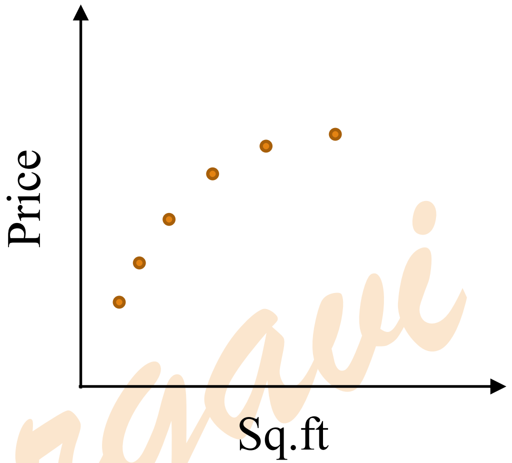
Bias Variance Trade-off



$$h(x) = w_0 + \ w_1 x$$

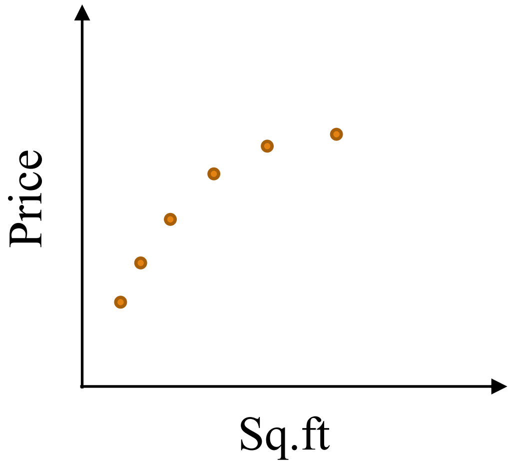
Underfit

High Bias



$$h(x) = w_0 + \ w_1 x + \ w_2 x^2$$

Correct Fit



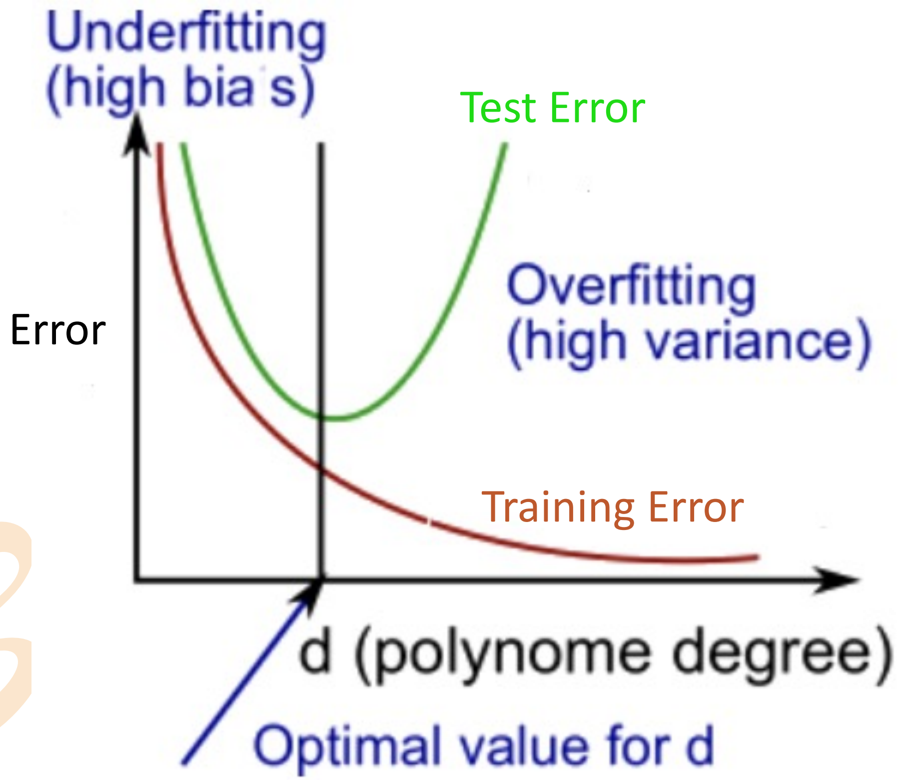
$$h(x) = w_0 + \ w_1 x + \ldots + w_{10} x^{10}$$

Overfit

High Variance

a

Bias Variance Trade-off (cont…)



Bias Variance Trade-off (cont…)

- **Getting more training examples:** Fixes high variance
- **Trying smaller sets of features:** Fixes high variance
- **Adding features:** Fixes high bias
- **Adding polynomial features:** Fixes high bias