

Regression Modeling

Training/Test Split

Training Set (Blue dots)

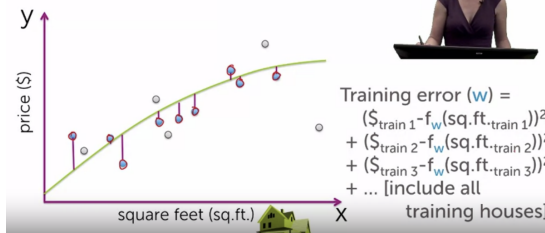
- The data selected to train the machine learning model

Test Set (Grey dots)

- Used after the model has been trained and validated to provide an unbiased evaluation of the model performance

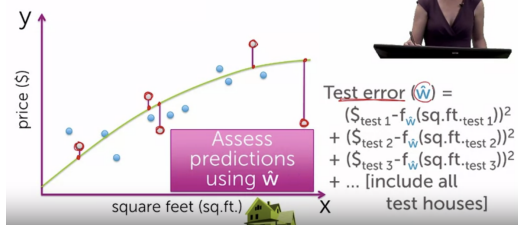
Training Error

Training error



Testing Error

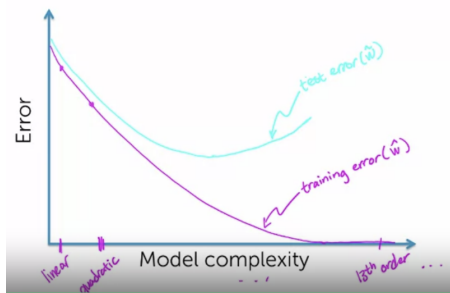
Test error



Training/Test Curves

Training error decreases with increasing model complexity as it fits the model better and better

Test error might decrease up until a certain complexity before the test error increases since higher model complexity means the line is more bespoke to the training data and less applicable to the test data.



ML Block Diagram

Start off with your data, then extract some features to model the predictions on. The ML model does a regression and outputs a predicted price.

The predicted price is measured against actual price and the error (residual squared) is fed into a ML algorithm that weighs the different features (via regression coefficients) and refines the ML model to better predict the house prices. This is iterative.

