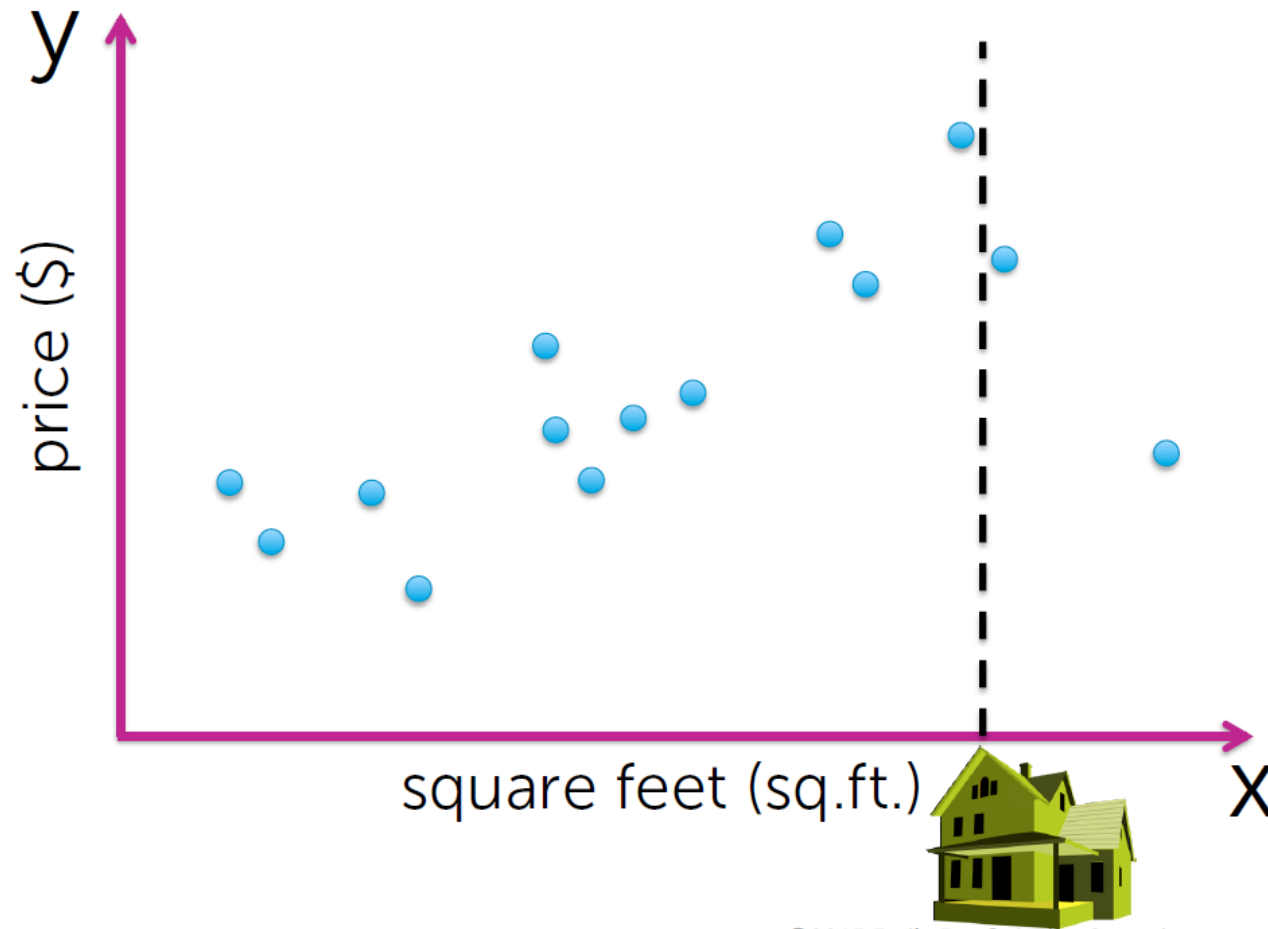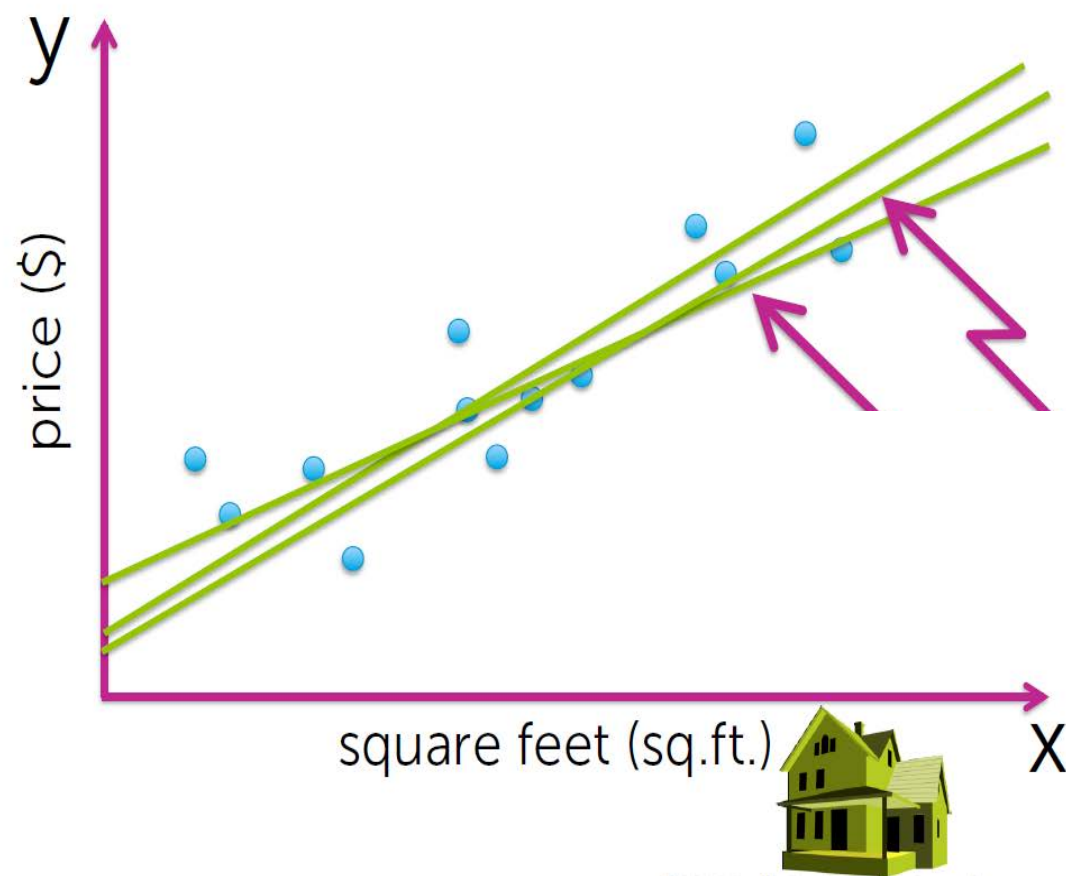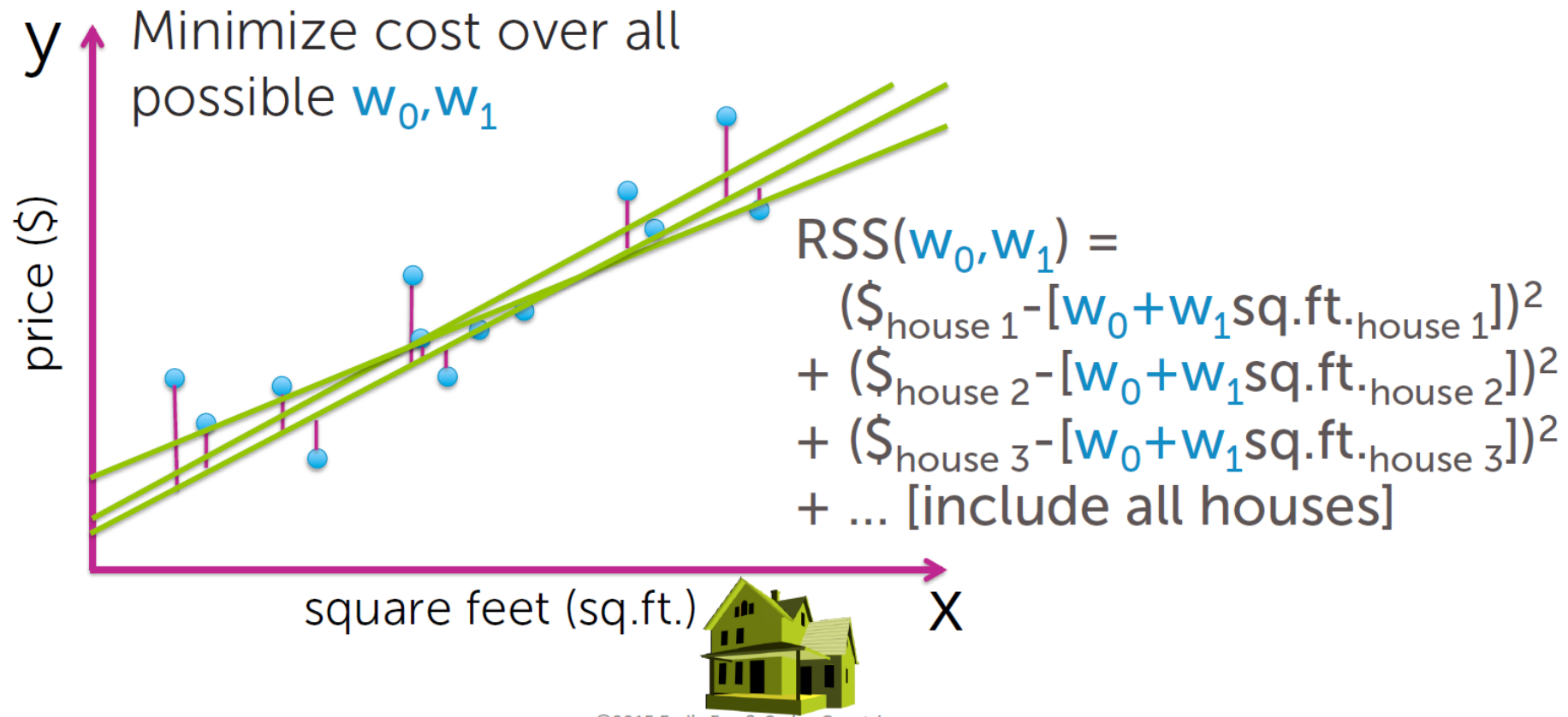# Linear Regression

Predicting house prices

Predict your house – By similar houses

Line of best fit ?

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

Residual Sum of Squares
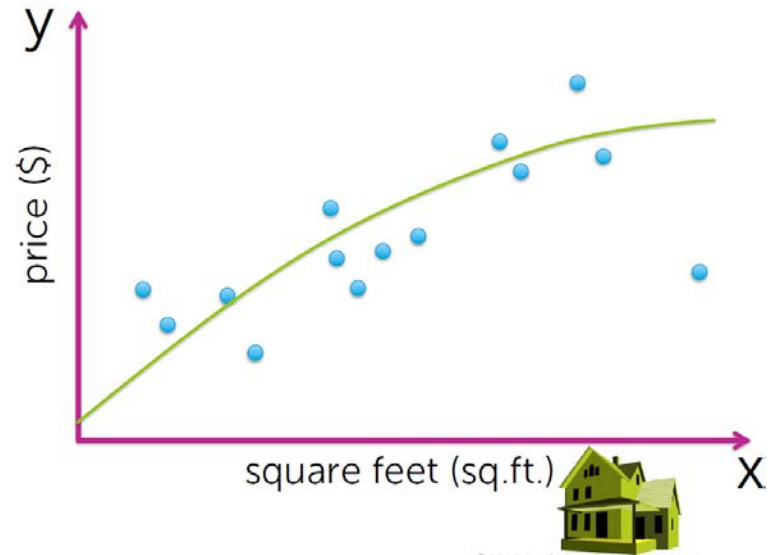

Minimize cost over all possible w_0, w_1

y

price ($)

square feet (sq.ft.)

x

$$RSS(w_0, w_1) =$$
$$(\$_{house\ 1} - [w_0 + w_1 sq.ft._{house\ 1}])^2$$
$$+ (\$_{house\ 2} - [w_0 + w_1 sq.ft._{house\ 2}])^2$$
$$+ (\$_{house\ 3} - [w_0 + w_1 sq.ft._{house\ 3}])^2$$
$$+ \dots \text{ [include all houses]}$$

Best fit ?

Bias- Variance Trade Off

Simple Rules
Better Generalization

Complex Rules
Better Performance of
Training data set

More Assumptions
More **Bias**
Simpler solutions

More Complexity
More **Variance**
Simpler solutions

UNDERFITTING

OVERFITTING

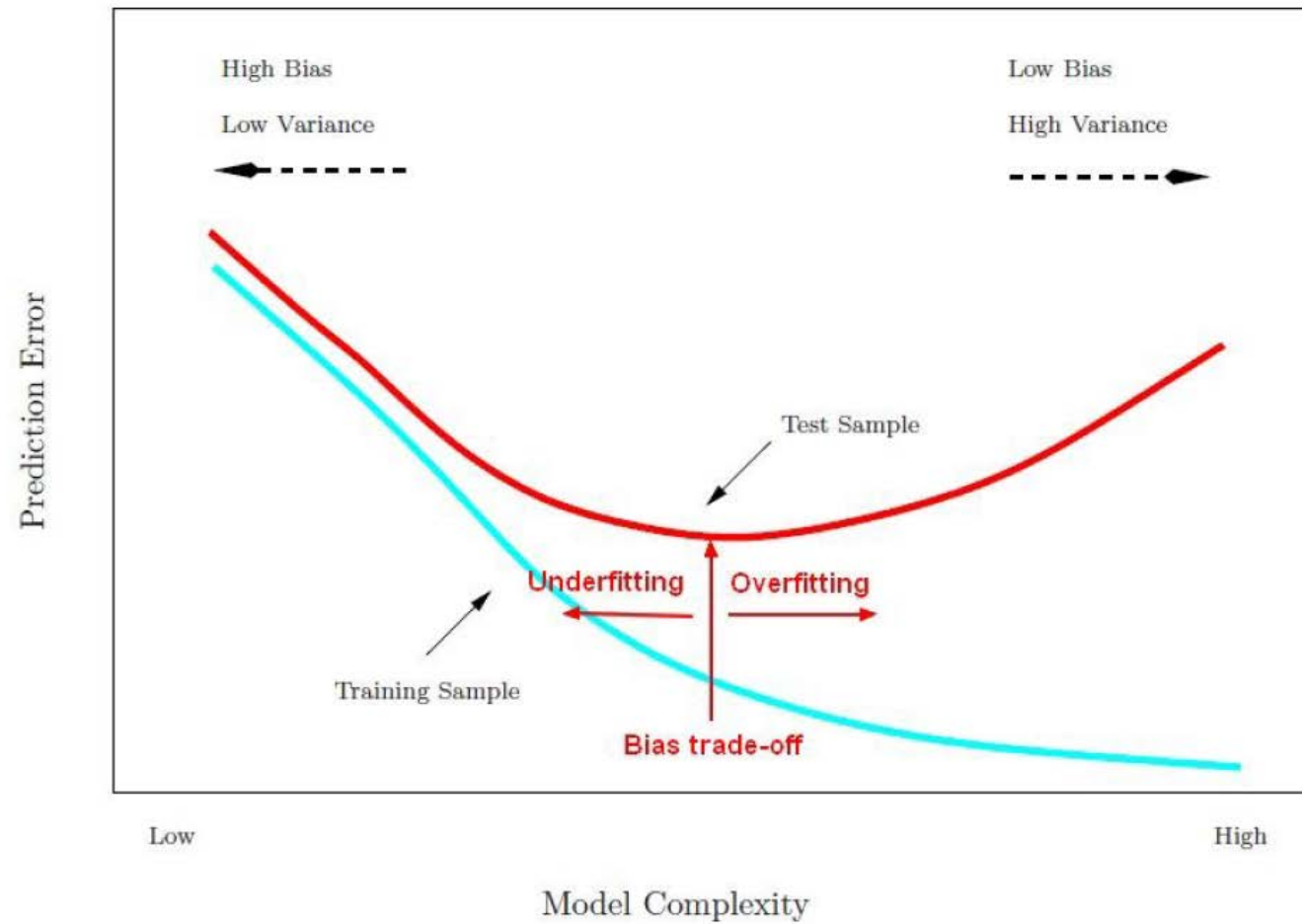Trade-off between our goals

Low Variance   High Variance

Low Bias

High Bias
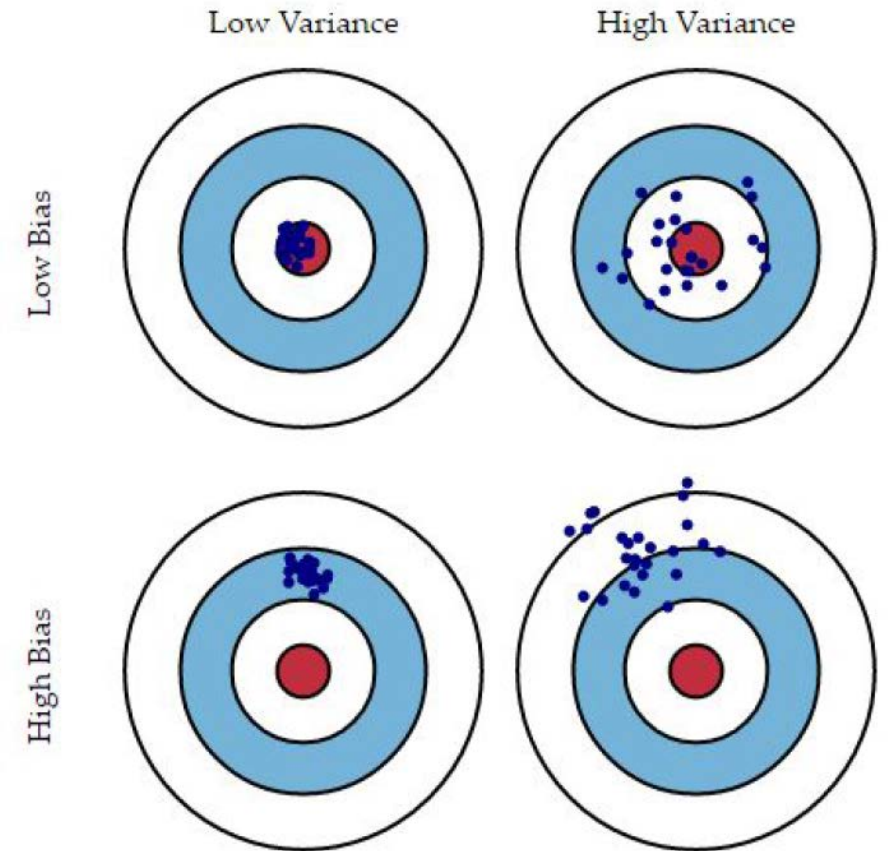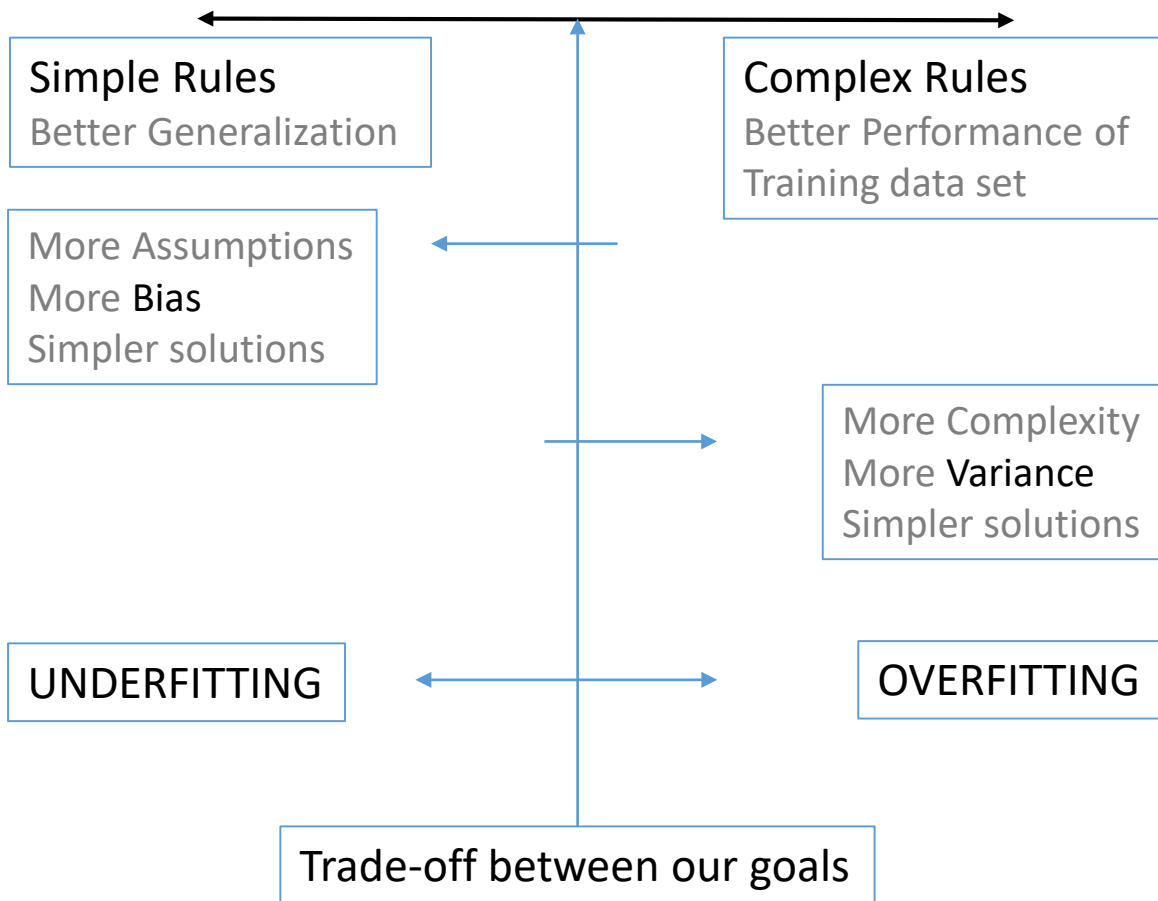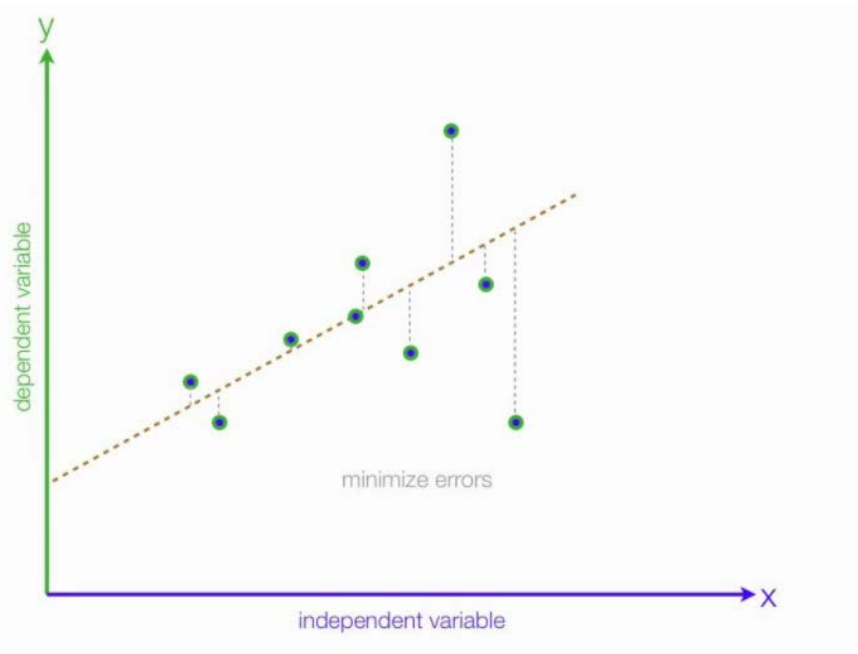
Bulls Eye - center of the target is a model that perfectly predicts the correct values.

Bias- Variance Trade Off

Terminology - 1

Terminology - 2

Correlation :
A statistic that measures the strength of the relationship between two variables.
From **correlation** we can only get an index describing the linear **relationship between** two variables

Regression :
We can predict the **relationship between** more than two variables and can use it to identify which variables x can predict the outcome variable y.

Covariance :
A measure of the tendency of two variables to vary together, **i.e** measure of **correlation**.
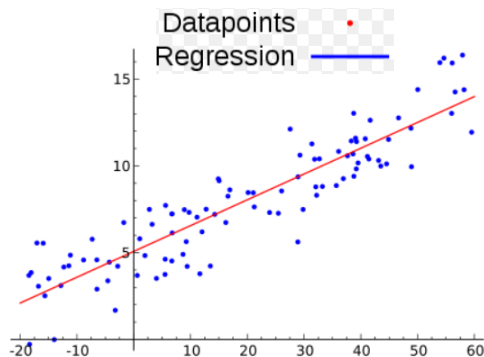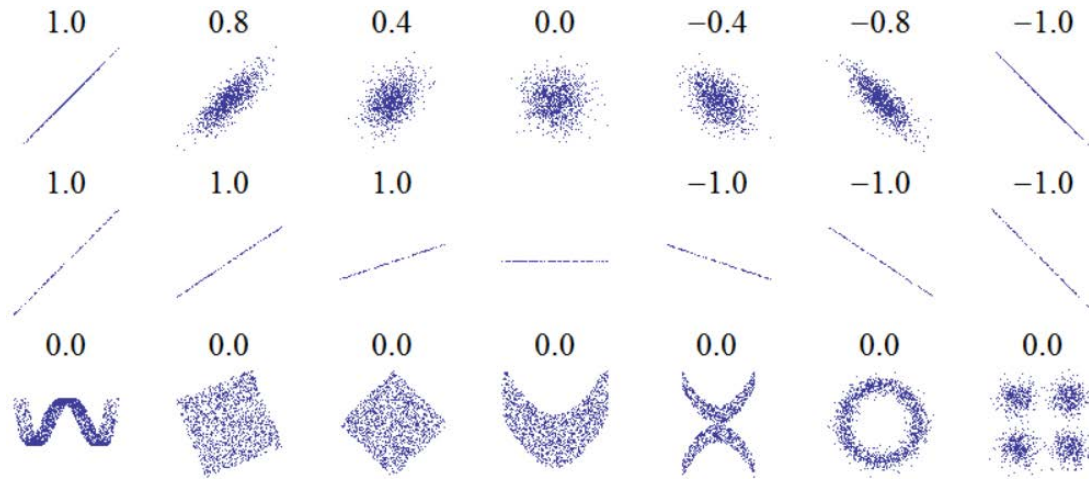
Formula to find the mean for X

$$\mu_x = \frac{\sum_{i=1}^{n} x_i}{n}$$

Formula to find the mean for Y

$$\mu_y = \frac{\sum_{i=1}^{n} y_i}{n}$$

Formula to find covariance of X & Y

$$cov(X,Y) = \frac{\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)}{(n-1)}$$

Linear Regression Examples

- Height of a person
- Stock Market predictions
- Add some more...