# Polynomial regression, Regularization and ridge regression

Linear regression

$$Y = \hat{\beta}_0 + \hat{\beta}_1 x + \varepsilon$$

Polynomial regression

$$Y = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 + \dots + \hat{\beta}_n x^n + \varepsilon$$

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x + \hat{\beta}_2 x^2 + \dots + \hat{\beta}_n x^n$$

# Linear regression on house pricing dataset

Demonstration in jupyter notebook

# What are Bias and Variance

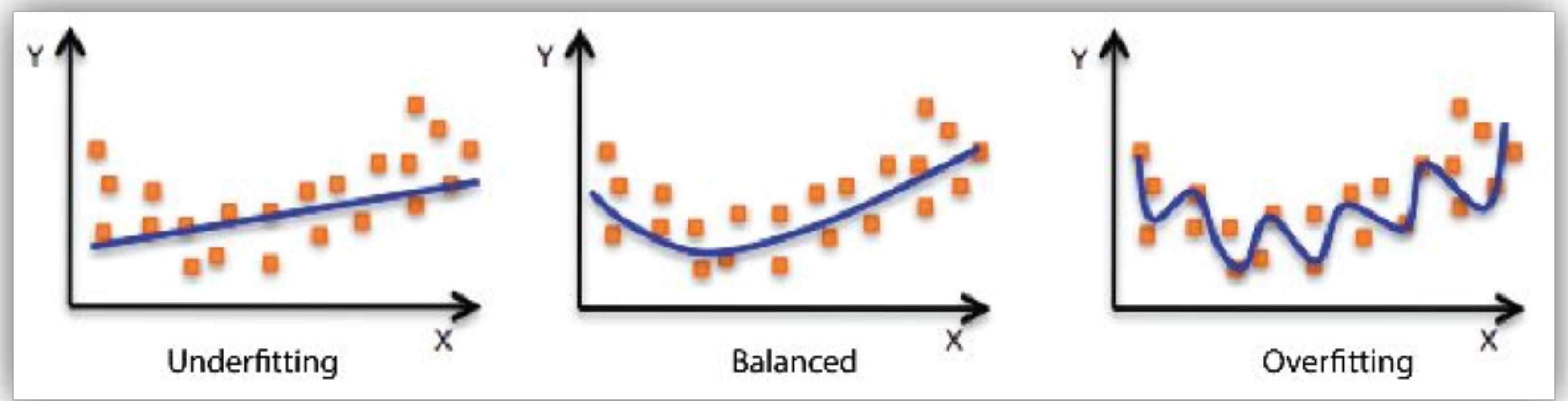
**What is Bias?**

- The amount our prediction differs  from the correct value that we are trying to predict
- High bias - Oversimplifies the model, high error on train and test sets

**What is Variance?**

- The amount our prediction will change given a different training data
- High variance - Learns too much on the training set, lower error on train and higher error on test

# Overfitting and Underfitting



# Bias-Variance tradeoff

- Model is too simple, has few parameters -> high bias, low variance
- Model is too complex, has a lot of parameters -> low bias, high variance

- Increasing bias will decrease variance (and vice versa)

# Probability VS Likelihood

- Probability -> how likely is it to get our desired outcome when the parameters are fixed?

- Likelihood -> How likely is it to get the values of the parameters when the outcome is fixed?

# Probability VS Likelihood

- We toss a coin 10 times, we know its a fair coin, what is the probability of getting heads every time?

- We toss a coin 10 times and we get 10 heads, what is the likelihood that the coin was fair?

# MSE at point x

$$\boxed{\begin{aligned} Var(\hat{f}(x)) &= E(\hat{f}(x)^2) - (E(\hat{f}(x)))^2 \\ Bias(\hat{f}(x)) &= E(\hat{f}(x)) - f(x) \end{aligned}}$$

$$MSE(x) = (f(x) - \hat{f}(x))^2$$

$$\boxed{E(MSE(x))} = E[(f(x) - \hat{f}(x))^2] = E[f(x)^2 - 2f(x)\hat{f}(x) + \hat{f}(x)^2] =$$

$$E(f(x)^2) - E(2f(x)\hat{f}(x)) + E(\hat{f}(x)^2) = f(x)^2 - 2f(x)E(\hat{f}(x)) + E(\hat{f}(x)^2) =$$

$$E(\hat{f}(x)^2) - (E(\hat{f}(x)))^2 + (E(\hat{f}(x)))^2 + f(x)^2 - 2f(x)E(\hat{f}(x)) =$$

$$\boxed{Var(\hat{f}(x)) + Bias^2(\hat{f}(x))}$$

# Regularization

Regularizations are techniques used to reduce the error and avoid overfitting.

**Our goal:** increase bias in a “nice” way to decrease the variance and avoid overfitting

Ridge and Lasso Regressions involve adding penalties to the regression function

# Ridge regression

$$\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

For Linear Regression

$$\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 + \lambda \sum_{i=1}^{k} \hat{\beta}_i^2$$

For Ridge Regression

# Lasso Regression

$$\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

For Linear Regression

$$\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 + \lambda \sum_{i=1}^{k}|\hat{\beta}_i|$$

For Lasso Regression

# Ridge regression continued

$$min_{\beta} \ u'u = (Y - X\hat{\beta})'(Y - X\hat{\beta})$$

For Linear Regression

$$min_{\beta} \ u'u = (Y - X\hat{\beta})'(Y - X\hat{\beta}) + \lambda\hat{\beta}'\hat{\beta}$$

For Ridge Regression

$$\frac{\partial}{\partial\beta}u'u = 2(X'X)\hat{\beta} - 2X'Y + 2\lambda\hat{\beta} = 0$$

$$2(X'X)\hat{\beta} + 2\lambda\hat{\beta} = 2X'Y$$

$$(X'X + \lambda I)\hat{\beta} = X'Y$$

$$\boxed{\widehat{\beta_{ridge}} = (X'X + \lambda I)^{-1}X'Y}$$

VS

$$\widehat{\beta_{linreg}} = (X'X)^{-1}X'Y$$

# Ridge regression on house pricing dataset

Demonstration in jupyter notebook