# Understanding the Bias-Variance Tradeoff

https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229

## What is bias? (Accuracy)

Bias is the difference between the average(expectation) prediction of our model and the ground truth.

$$E(\hat{y} - y)$$

## What is Variance? (Precision)

Variance is the variability of model prediction.

$$Var(\hat{y} - \bar{y})$$

## Mathematically

Suppose we have the variable

$$y = f(x) + e$$

Where $e$ is the error term and it's normally distributed with a mean of 0, and we have

$$error = E\left(Y - \widehat{f(x)}\right) = E(y - \hat{y})$$

Since we want to eliminate the effects of direction of errors (positive or negative), we use expected squared error to measure the fitness of model, which equals to

$$E(error^2) = E[(y - \hat{y})^2]$$

Since $Var(X) = E(X^2) - E(X)^2$, So we have

$$E(error^2) = E[(y - \hat{y})^2] = Var(y - \hat{y}) + E(y - \hat{y})^2$$

The first item $Var(y - \hat{y})$ is called variance, and the left one is called bias.