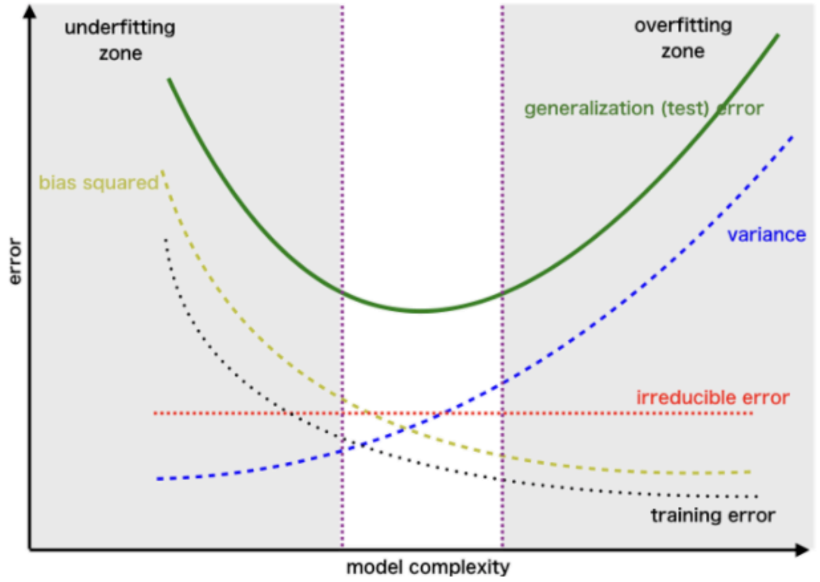
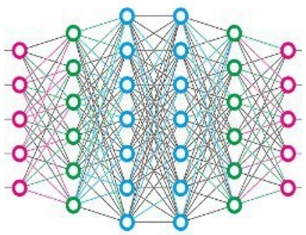
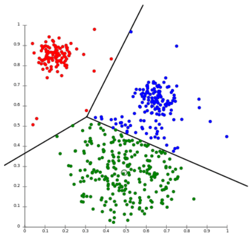
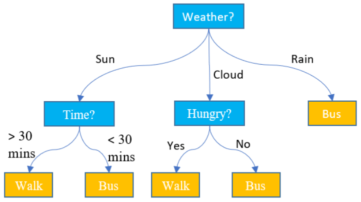
CS 445
Bias and Variance of Models



Instructor: Dr. Kevin Molloy

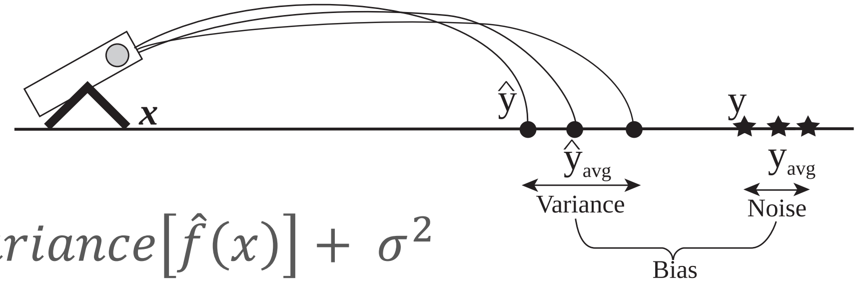






Bias-Variance Decomposition

Problem of reaching a target y by firing projectiles from x (regression problem)



$$E\left[\left(y - \hat{f}(x)\right)^2\right] = \left(Bias\left[\hat{f}(x)\right]\right)^2 + Variance\left[\hat{f}(x)\right] + \sigma^2$$

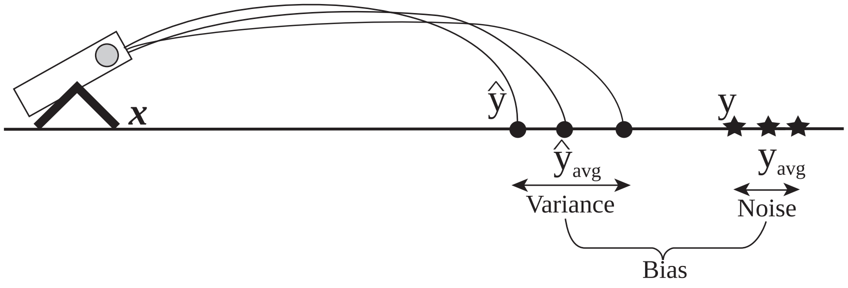
y_{avg} is the average of the outcomes that occur when applying the same power and launch angle (*x*) (yet the projectile ends up in a slightly different location). The difference in outcomes is the result of **noise**. We can not control or eliminate this uncertainty.

$\hat{y}_{avg}$ is the average *prediction* across a **set** of models ***M***. Each of these models may be different for many reasons (different training data, different method, etc.). This is known as the **variance**.

Figures from Tan, Steinbach, Karpatne, Kumar. *Intro to Data Mining.*

Bias-Variance Decomposition

Problem of reaching a target y by firing projectiles from x (regression problem)

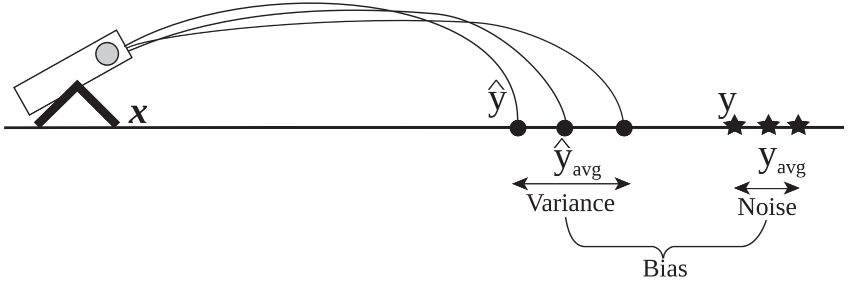


$$Expected\ Error = \left[\left(y - \hat{f}(x)\right)^2\right] = \left(Bias\left[\hat{f}(x)\right]\right)^2 + Variance\left[\hat{f}(x)\right] + \sigma^2$$

σ^2 is the variance in the **noise** term, shown as y_{avg} in the diagram. When applying the same power and launch angle (*x*), noise is the variable that accounts for how the projectile can end up in a slightly different location. We can not control or eliminate this uncertainty.

Figures from Tan, Steinbach, Karpatne, Kumar. *Intro to Data Mining.*

Bias-Variance Decomposition



Problem of reaching a target y by firing projectiles from x (regression problem)

$$Expected\ Error = \left[\left(y - \hat{f}(x)\right)^2\right] = \left(Bias\left[\hat{f}(x)\right]\right)^2 + Variance\left[\hat{f}(x)\right] + \sigma^2$$

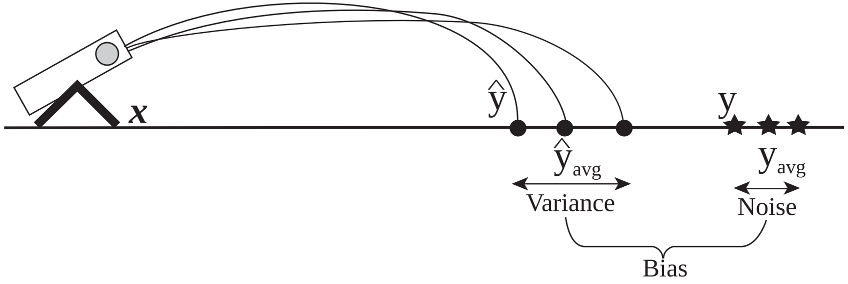
$$Bias\left[\hat{f}(x)\right] = E[(\hat{f}(x)] - f(x)$$

The difference between y_{avg} *and* $\hat{y}_{avg}$ is known as the **bias**.

Example: The simplest model would make the same predictions regardless of the input. In this case, the difference between y_{avg} and $\hat{y}_{avg}$ would be large (**high bias**) (since the variance term would be very small).

Figures from Tan, Steinbach, Karpatne, Kumar. *Intro to Data Mining.*

Bias-Variance Decomposition



Problem of reaching a target y by firing projectiles from x (regression problem)

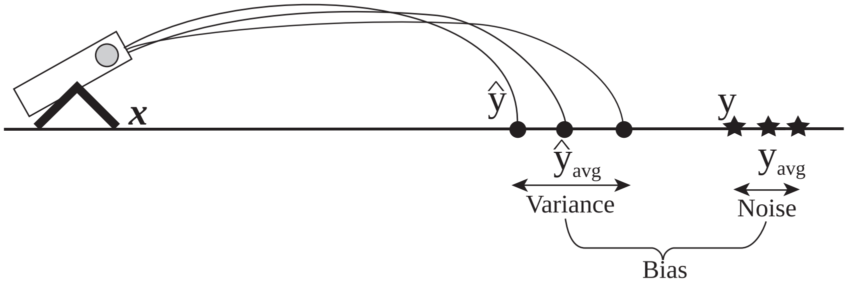
$$Expected\ Error = \left[\left(y - \hat{f}(x)\right)^2\right] = \left(Bias\left[\hat{f}(x)\right]\right)^2 + Variance\left[\hat{f}(x)\right] + \sigma^2$$

$\hat{y}_{avg}$ is the variance of the average *prediction* across a **set** of models ***M***. Each of these models may be different for many reasons (different training data, different model, etc.). This is known as the **variance**. Think of it as the if you built 10 models (maybe with different hyperparameters or different training data), the predictions would differ. This is the "variance" in the models (or the models' predictions).

Figures from Tan, Steinbach, Karpatne, Kumar. *Intro to Data Mining.*

Bias-Variance Decomposition

Problem of reaching a target y by firing projectiles from x (regression problem)

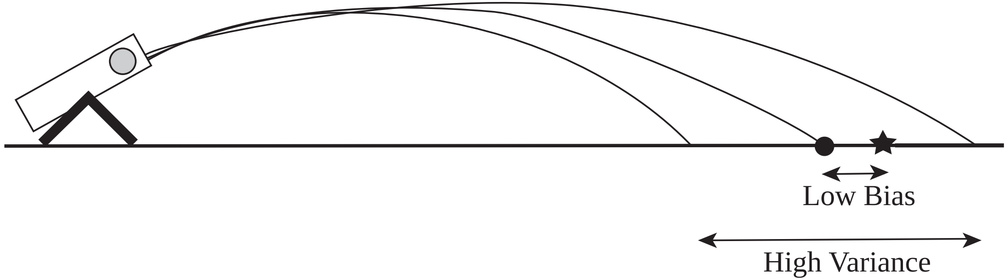


Example 2: The model changes drastically with the change of training data (it is overfitting). This means that the spread of $\hat{y}_{avg}$ values (changes per model/set of training data) is high, this is known as **high variance**.

Figures from Tan, Steinbach, Karpatne, Kumar. *Intro to Data Mining.*

Bias-Variance Decomposition

Overfitting



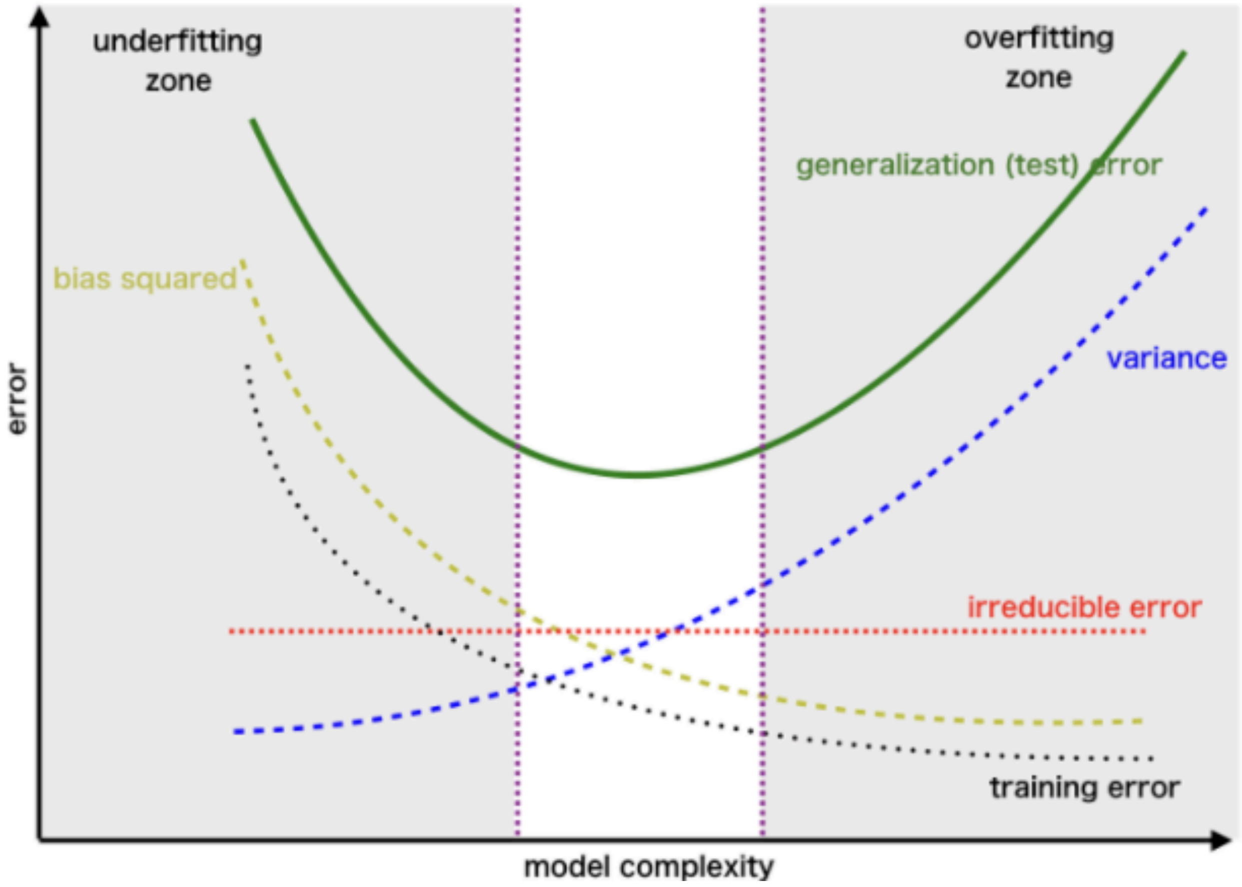
Gen error for a classification model m can be given by:

$$genError(m) = (c_1 \times noise) + bias(m) + (c_2 \times noise)$$

Where c1 and c2 are constants. The idea of bias-variance is easier explained with regression, and I will not be testing you on bias-variance decomposition with classification, however, you should be able to identify classification models that exhibit low/high bias or low/high variance.

Figures from Tan, Steinbach, Karpatne, Kumar. *Intro to Data Mining.*

Bias-Variance Decomposition



Bias-Variance Decomposition