Cornell Bias Variance Tradeoff

$P(x, y) = P(y|x)\, P(x)$

$\bar{y}(x) = E_{y|x}(y) = \int y\, P(y|x)\, dy$

Expected lable

given x, we need to find y — — Algorithm

D = data set $\longrightarrow A(D)$ that takes data D

h_D = outcomes of ML algorithm

Expected Error (test) given h_D

Using Squared loss. $\longrightarrow$ for regression its easy loss hence

$$\boxed{E\left[\left[h_D(x) - y\right]^2\right]}$$

generalization error given h_D

$$= \int_x \int_y (h_D(x) - y)^2 \, dy\, dx\; P(x, y)$$

Expected Classifier $= \bar{h} = E_{D \sim P^n}\left[A(D)\right] = \int_D h_D\, P(D)\, dD$

Expected Error of A :-

$$E_{\substack{(x,y) \sim P \\ D \sim P^n}} \left[(h_D(x) - y)^2\right]$$

$$= \int_D \int_x \int_y (h_D(x) - y)^2\, P(x, y)\, P(D)\, dy\, dx\, dD$$

Decomposition

$$\Rightarrow E\left[(h_D(x)-y)^2\right] = E_{x,y,D}\left[\left[h_D(x) - \bar{h}(x) + \bar{h}(x) - y\right]^2\right]$$

$$= E\left[\left[\left(h_D(x) - \bar{h}(x)\right) + \left(\bar{h}(x) - y\right)\right]^2\right]$$

$$(A + B)^2$$

$$= E\left[\left(h_D(x) - \bar{h}(x)\right)^2\right] + E\left[\left(\bar{h}(x) - y\right)^2\right] + 2\left[\left(h_D(x) - \bar{h}(x)\right)\left(\bar{h}(x) - y\right)\right]$$

$$= A^2 \qquad B^2 \qquad 2AB$$

$$= 0$$

$$= E_D\left(h_D(x)\right) - \bar{h}(x)$$

$$= \bar{h}(x) - \bar{h}(x)$$

$$= 0$$

$$\boxed{E(A) = E\left[\left(h_D(x) - \bar{h}(x)\right)^2\right] + E\left[\left(\bar{h}(x) - y\right)^2\right]}$$

label

Avg predictio

— mean for value

random variable

=

Variance

bias

2nd term

Add and subtract mean $\bar{y}$

expected value of y

$$E_{x,y}\left[\left[\bar{h}(x) - y\right]^2\right] =$$

$$= E_{x,y}\left[\left[\underbrace{\left(\bar{h}(x) - \bar{y}(x)\right)}_{a} + \underbrace{\left(\bar{y}(x) - y\right)}_{b}\right]^2\right]$$

$$= \underbrace{E\left[\left(\bar{h}(x) - \bar{y}(x)\right)^2\right]}_{a^2} + \underbrace{E\left[\left(\bar{y}(x) - y\right)^2\right]}_{b^2} + \underbrace{2\left[\bar{h}(x) - \bar{y}(x)\right]\left[\bar{y}(x) - y\right]}_{2ab}$$

$$= E\left[\left[\bar{h}(x) - \bar{y}(x)\right]^2\right] + E\left[\left(\bar{y}(x) - y\right)^2\right] + 0$$

Same argument as before

MSE =

$$\underbrace{E\left(\bar{h}(x) - \bar{y}(x)\right)^2}_{\text{Bias}^2} + \underbrace{E\left[\left(\bar{y}(x) - y\right)^2\right.}_{\text{Noise}} + \underbrace{E\left[h_D(x) - \bar{h}(x)^2\right]}_{\text{Variance}}$$

→ finally it decomposes to these three values

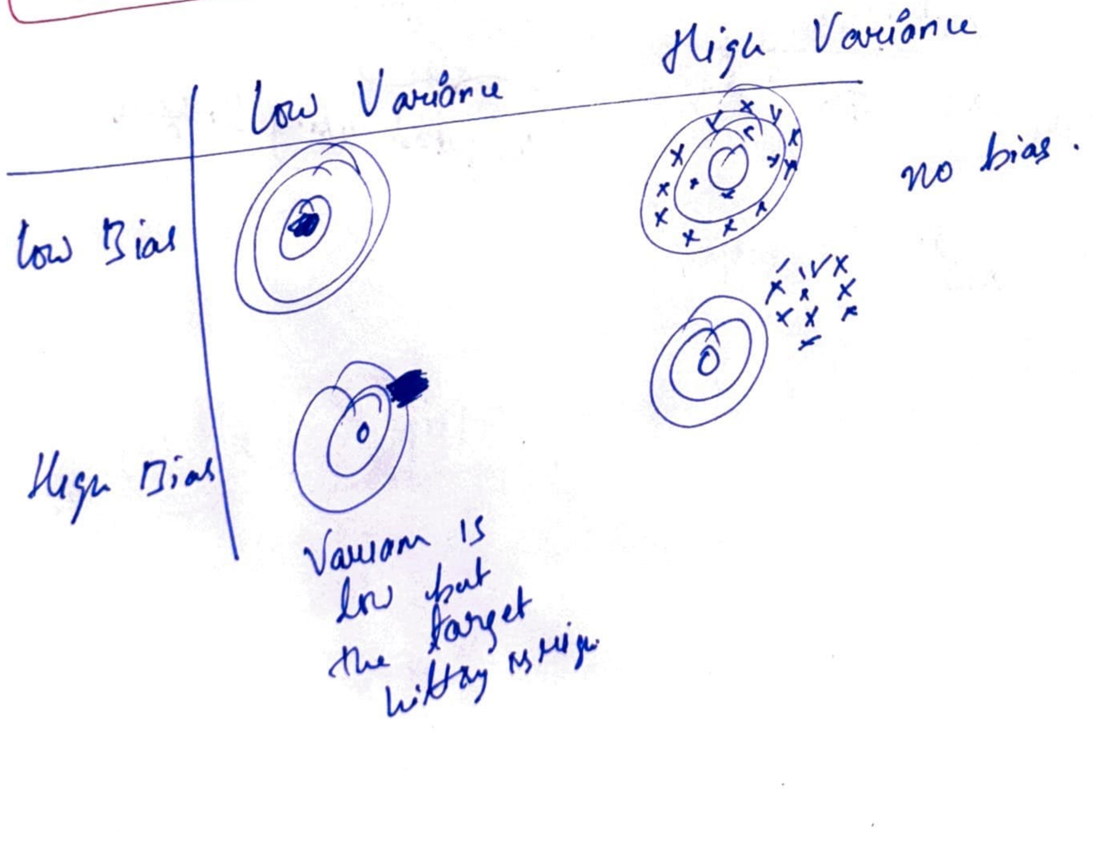
Noise does'nt matter.

How much error would I still get ?.

→ Bias says this

→ always make some mistake as data is biased over something.

$\boxed{\text{Ultimate error = Bias + Variance + Noise}}$



L2 Regularization $= \text{loss} + \frac{1}{\alpha} \alpha \|w\|^2$

Size = (slope × weight) + y Intercept

Size = (slope 1 × weight) + (slope 2 × Age) + y-Intercept

Newtons Method :-

$$w^{k+1} = w^k - \eta \underbrace{H_E}_{} (w^k)^{-1} \nabla_w E(w^k)^T \quad ; \eta = 1$$

Hession — very difficult to calculate
Even harder to Invert

learning rate issue

~~Hession~~ × S.O. Derivative

Quick Prop

$$w^{k+1} = w^k - \left[\frac{E'(w^k) - E'(w^{k-1})}{\Delta w^{k-1}} \right]^{-1} E'(w^k)$$

$$\Delta w^k = \frac{\Delta w^{k-1}}{Err(w^k) - Err'(w^{k-1})} \underbrace{Err'(w^k)}_{}$$

computed using Backprop

$$w^{k+1} = w^k - \Delta w^k$$