CS4780 L19 — Bias Variance Tradeoff

$D = \{(x_1, y_1) \dots (x_n, y_n)\}$ $y_i \in \mathbb{R}$

Each (x, y) iid

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

y has distribution for same x

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

Eg. Same house could be sold for 50,000 – 70,000\$.

Classifier being learned h_D

$$h_D = A(D)$$

A – Algorithm say SVM, Perceptron

Expected test error given h_D

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

Simplicity we pick square loss.

$$= \int_x \int_y \left[h_D(x) - y\right]^2 P(x,y) \, dy \, dx$$

Now, h_D is also a random variable set. As for different D, h_D changes. Expected value of h_D

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

$\bar{h}$ – Average classifier on infinitely many datasets.

Expected Error of A

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

Take D of n datapts. from P. Train to get $h_D(x)$, take a (x,y) test point and get $(h_D(x) - y)^2$

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

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

Add and subtract $\bar{h}(x)$

$$= \underbrace{E_{x,D}\left[\left[h_D(x) - \bar{h}(x)\right]^2\right]}_{a^2} + \underbrace{E_{x,y}\left[\left(\bar{h}(x) - y\right)^2\right]}_{b^2} + 2\underbrace{E_{\substack{x,y \\ D}}\left[\left(h_D(x) - \bar{h}(x)\right)\left(\bar{h}(x) - y\right)\right]}_{2ab \,\to\, 0}$$

(see next page)

$2ab = 0$

$$E_{x,y}\left[E_D[h_D(x) - \bar{h}(x)]\right]$$

because of linearity of Expected value

$$E_D[h_D(x) - \bar{h}(x)$$

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

$$= 0$$

$E_D[\bar{h}(x)] = \bar{h}(x)$

D is not present here

Thus, Expected Error of A remains as

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

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

add & subtract $\bar{y}(x)$

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

Since $E_{xy}[(\bar{y}(x) - y)] = 0$ (see above explanation & below)

$$E_x\left[E_{y|x}[\bar{y}(x) - y]\,(\bar{h}(x) - \bar{y}(x))\right]$$

[To calculate E_{xy} first we calculate E_x and then E of y given x "$E_{y|x}$"]

so, $\bar{y}(x) - E_{y|x}[y] = \bar{y}(x) - \bar{y}(x)$

It's a constant as it's a mean.

We are ~~...~~ left with

Expected $\overset{\text{Test}}{\text{Error}}$ of A

Maybe add more features

$$= \underset{x,d}{E}\Big[\underbrace{\big[h_D(x) - \bar{h}(x)\big]^2}\Big] + \underset{x,y}{E}\Big[\underbrace{\big[\bar{h}(x) - \bar{y}(x)\big]^2}\Big] + E\Big[\underbrace{(\bar{y}(x) - y)^2}\Big]$$

Variance of classifier

how much the classifier $h_D(x)$ varies w.r.t to $\bar{h}(x)$ ie average/expected classifier

BIAS^2

If have unlimited data and could get expected classifier $\bar{h}(x)$ how much error can I get to predict the expected label. (not a particular y but average y)

If data is non linear and I use linear classifier, no matter how much data I have I will always have high bias.

NOISE of data