Bias-Variance Tradeoff

DT $\rightarrow$ $\ddot{\alpha}$

$$D = \{(x_1, y_1), \dots, (x_n, y_n)\} \sim P(X, Y)$$

x_i: d-dim., y_i: $\mathbb{R}$

Expected Label (given x)

$$\bar{y}(x) = \mathbb{E}_{y|x}[Y] = \int_y y \Pr(y|x)\,dy$$

learning alg: $\mathcal{A}$
classifier: h_D

Expected Test Error (given h_D)

$$\mathbb{E}_{(x,y)\sim P}\left[(h_D(x) - y)^2\right] = \int_x \int_y (h_D(x) - y)^2 \Pr(x,y)\,dx\,dy$$

$D \sim P(X,Y) \longrightarrow$ R.V.
$h_D \longrightarrow$ R.V.

Expected Classifier (given $\mathcal{A}$)

$$\bar{h} = \mathbb{E}_{D\sim P}[h_D] = \int_D h_D \Pr(D)\,dD$$

Expected Test Error (given $\mathcal{A}$)

$$\mathbb{E}_{\substack{(x,y)\sim P \\ D\sim P}}\left[(h_D(x) - y)^2\right] = \int_D \int_x \int_y (h_D(x) - y)^2 \Pr(x,y)\Pr(D)\,dx\,dy\,dD$$

Decomposition of Expected Test Error

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

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

$$= 2\,\mathbb{E}_{x,y}\left[\mathbb{E}_D[h_D(x) - \bar{h}(x)](\bar{h}(x) - y)\right]$$

$$= 2\,\mathbb{E}_{x,y}\left[(\underbrace{\cancel{\mathbb{E}[h_D(x)]}}_{\bar{h}} - \bar{h}(x))(\bar{h}(x) - y)\right]$$

0

$$= 0$$

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

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

$$= \underbrace{\mathbb{E}_{x,y}\left[(\bar{y}(x) - y)^2\right]}_{\text{Noise}} + \underbrace{\mathbb{E}_x\left[(\bar{h}(x) - \bar{y}(x))^2\right]}_{\text{Bias}} + 2\,\mathbb{E}_{x,y}\left[(\bar{h}(x) - \bar{y}(x))(\bar{y}(x) - y)\right]$$

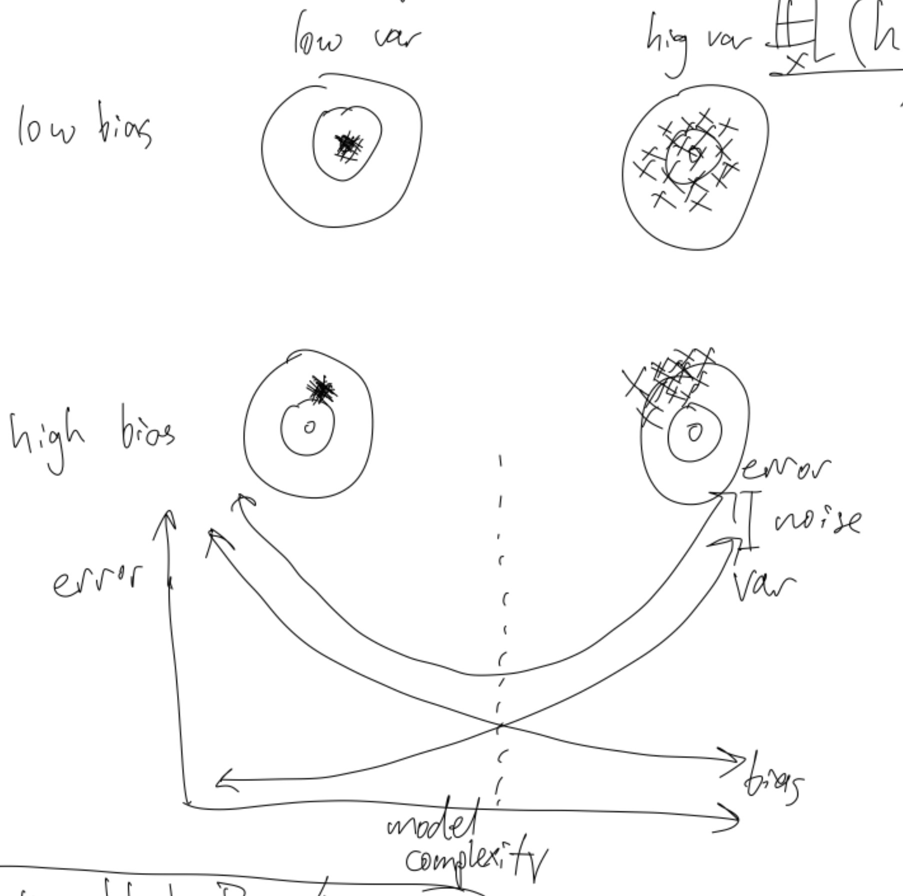
$$= \mathbb{E}_x\left[\mathbb{E}_{y|x}[\bar{y}(x) - y](\bar{h}(x) - \bar{y}(x))\right]$$

$$= \mathbb{E}_x\left[(\bar{y}(x) - \underbrace{\cancel{\mathbb{E}_{y|x}[y]}}_{\bar{y}(x)})(\bar{h}(x) - \bar{y}(x))\right]$$

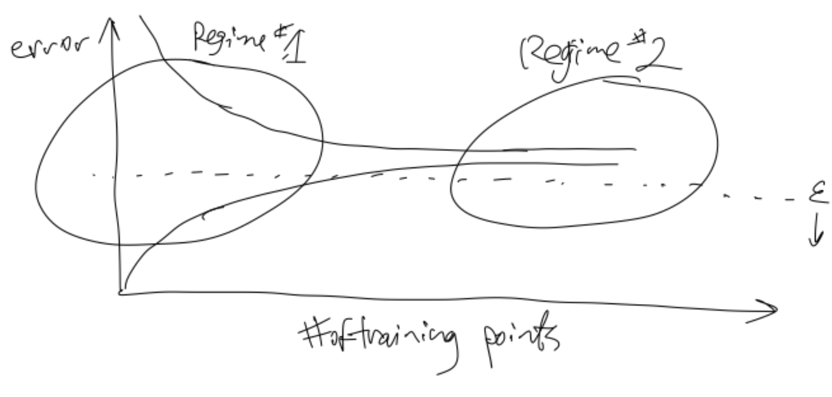
0

$$= 0$$

$$\mathbb{E}_{x,y,D}\left[(h_D(x) - y)^2\right] = \underbrace{\mathbb{E}_{x,D}\left[(h_D(x) - \bar{h}(x))^2\right]}_{\text{Variance}} + \underbrace{\mathbb{E}_{x,y}\left[(\bar{y}(x) - y)^2\right]}_{\text{noise}} + \underbrace{\mathbb{E}_x\left[(\bar{h}(x) - \bar{y}(x))^2\right]}_{\text{bias}}$$



Detecting High Bias/Variance



1. High Variance
   - train error is lower than ε
   - test error is higher than ε

   Treatments
   1. Add more data
   2. Reduce model complexity
   3. Bagging
2. High Bias
   - train error is higher than ε

   Treatments
   1. Use more complex model
   2. Add features
   3. Boosting