Announcements

- Homework 6 is due Tuesday
- My office hours tomorrow will be 11 AM - 12 PM.
- Email is coming for vote on midterm ground rules
- Project Assignment will be posted soon

Today's Lecture

- Other error measures and target types
- Approximation-generalization tradeoff
  - Bias Variance decomposition
  - Learning curves

ERROR MEASURES AND TARGET TYPES

WE HAD $E_{in}(h) = \frac{1}{N} \sum_{n=1}^{N} [\![h(\underline{x}_n) \neq f(\underline{x}_n)]\!]$

$= \text{IN-SAMPLE ERROR RATE.}$

$$E_{out}(h) = P[h(\underline{x}) \neq f(\underline{x})]$$

CAN INSTEAD USE:

$$E_{in}(h) = \frac{1}{N} \sum_{n=1}^{N} \left[h(\underline{x}_n) - f(\underline{x}_n) \right]^2$$

$$E_{out}(h) = \mathbb{E}\{[h(\underline{x}) - f(\underline{x})]^2\}$$

($\mathbb{E}$ IS W.R.T. $p(\underline{x})$).

here: can have $h \in \mathbb{R}$, $f \in \mathbb{R}$.

FOR $h, f \in \mathbb{R}$, WE WILL USE ~~BA~~ BIAS AND VARIANCE TO ASSESS COMPLEXITY TRADEOFFS.

APPROXIMATION – GENERALIZATION TRADEOFF

→ USE BIAS AND VARIANCE.

NOTE: BELOW, h_g IS BEST HYPOTHESIS AS CHOSEN BY $\mathcal{D}$ (USING LEARNING ALGORITHM $\mathcal{A}$.)

target fn.

$$E_{OUT}\left(h_g^{(\mathcal{D})}\right) = \mathbb{E}_{\underline{x}}\left\{\left[h_g^{(\mathcal{D})}(\underline{x}) - f(\underline{x})\right]^2\right\}$$

TAKE $\mathbb{E}_{\mathcal{D}}$ OF BOTH SIDES

w.r.t. all datasets $\mathcal{D}$ of size N.

$$\mathbb{E}_{\mathcal{D}}\left\{E_{OUT}\left(h_g^{(\mathcal{D})}\right)\right\} = \mathbb{E}_{\mathcal{D}}\left\{RHS\right\}$$

DEFINE: $$\bar{h}_g(\underline{x}) = \mathbb{E}_{\mathcal{D}}\left\{h_g^{(\mathcal{D})}(\underline{x})\right\}$$

$$\approx \frac{1}{K}\sum_{k=1}^{K} \underbrace{h_{gk}(\underline{x})}_{\text{BEST HYPOTH. BASED ON } \mathcal{D}_k.}$$

$$\mathbb{E}_{\mathcal{D}}\left\{E_{OUT}\left(h_g^{(\mathcal{D})}\right)\right\}$$

$$= \mathbb{E}_{\underline{x}}\left\{\underbrace{\mathbb{E}_{\mathcal{D}}\left\{\left[h_g^{(\mathcal{D})}(\underline{x}) - \bar{h}_g(\underline{x})\right]^2\right\}}_{\text{var}(\underline{x})} + \underbrace{\left[\bar{h}_g(\underline{x}) - f(\underline{x})\right]^2}_{\text{bias}(\underline{x})}\right\}$$

(sometimes called $\text{bias}^2(\underline{x})$)

$$= \underbrace{\mathbb{E}_{\underline{x}}\{\text{var}(\underline{x})\}}_{\text{var}} + \underbrace{\mathbb{E}_{\underline{x}}\{\text{bias}(\underline{x})\}}_{\text{bias}}$$

(or bias^2)

$$\mathbb{E}_{\mathcal{D}}\left\{E_{OUT}\left(h_g^{(\mathcal{D})}\right)\right\} = \text{bias} + \text{var}.$$

AML EXAMPLE 2.8

$$f(x) = \sin(\pi x)$$

x IS SAMPLED UNIFORMLY ON $[-1, +1]$.

DATASET $\mathcal{D} = \{(x_1, y_1); (x_2, y_2)\}$

$\Rightarrow N = 2$.

$\mathcal{H}_0$: SET OF ALL HORIZONTAL LINES:

$$h_b(x) = b.$$

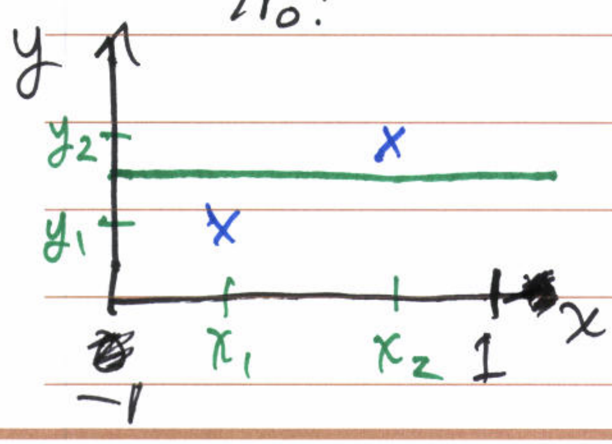
$\mathcal{H}_1$: SET OF ALL LINES:

$$h_{ab}(x) = ax + b$$

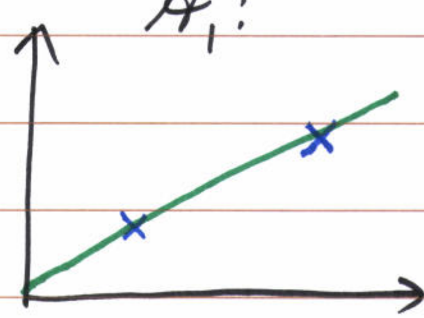
LEARNING ALGORITHM: $\mathcal{H}_0$: $b = \frac{y_1 + y_2}{2}$

$\mathcal{H}_1$: LINE PASSING THROUGH (x_1, y_1) AND (x_2, y_2).

$\mathcal{H}_0$:



$\mathcal{H}_1$:



LEARNING CURVES

REVIEW:

- BIAS-VAR. VIEWPOINT —

$$\mathbb{E}_{\mathcal{D}}\left\{E_{OUT}\left(h_g^{(\mathcal{D})}\right)\right\} = \mathbb{E}_{\underline{x}}\left\{\text{bias}(\underline{x}) + \text{var}(\underline{x})\right\}$$

$$= \text{bias} + \text{var}$$

$$\text{bias}(\underline{x}) = \left[\bar{h}_g(\underline{x}) - f(\underline{x})\right]^2$$

$$\text{var}(\underline{x}) = \mathbb{E}_{\mathcal{D}}\left\{\left[h_g^{(\mathcal{D})}(\underline{x}) - \bar{h}_g(\underline{x})\right]^2\right\}$$

- VC VIEWPOINT —

$$E_{OUT}\left(h_g^{(\mathcal{D})}\right) \leq E_{IN}\left(h_g^{(\mathcal{D})}\right) + \varepsilon(N, \mathcal{H}, \delta)$$

WITH PROBABILITY $\geq 1 - \delta$.

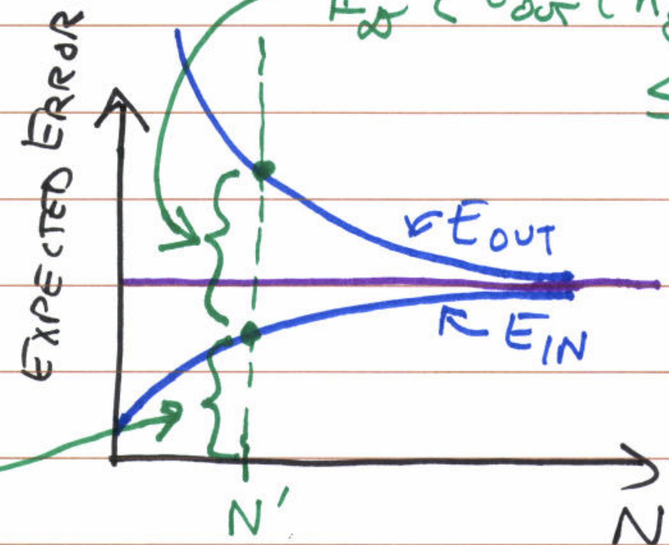
$$\mathbb{E}_{\mathcal{D}}\left\{E_{OUT}\left(h_g^{(\mathcal{D})}\right)\right\} \leq \mathbb{E}_{\mathcal{D}}\left\{E_{IN}\left(h_g^{(\mathcal{D})}\right)\right\} + \varepsilon(N, \mathcal{H}, \delta)$$

$$\varepsilon(N, \mathcal{H}, \delta) \leq \sqrt{\frac{8}{N} \ln \frac{4\left[(2N)^{d_{VC}} + 1\right]}{\delta}}$$

$$(\text{“}\varepsilon_{VC})$$

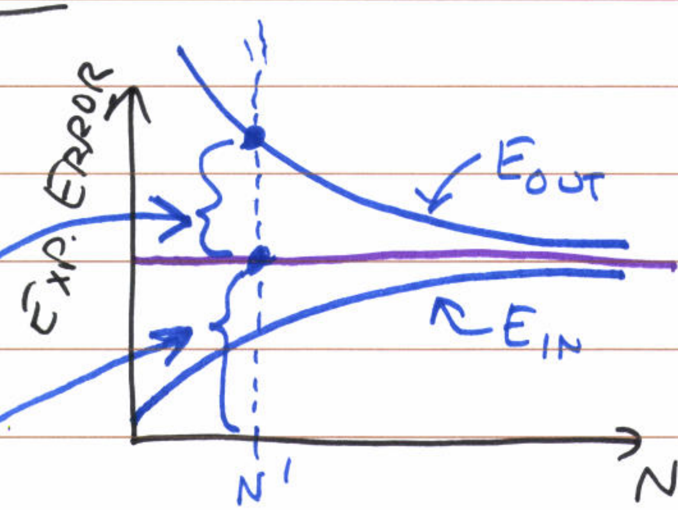
VC:

$$\mathbb{E}_{\mathcal{D}}\left\{ E_{OUT}(h_g^{(\mathcal{D})}) - E_{IN}(h_g^{(\mathcal{D})}) \right\} \le \mathcal{E}(N, \mathcal{H}, \delta)$$



$$\mathbb{E}_{\mathcal{D}}\left\{ E_{IN}(h_g^{(\mathcal{D})}) \right\}$$

BIAS-VAR



$$BIAS = \mathbb{E}_{x}\left\{ \left[\bar{h}_g(x) - f(x) \right]^2 \right\}$$

(BY ASSUMPTION)

$$\Rightarrow \quad VAR = \mathbb{E}_{x}\, \mathbb{E}_{\mathcal{D}}\left\{ \left[h_g^{(\mathcal{D})}(x) - \bar{h}_g(x) \right]^2 \right\}$$