▶ VC-Dimension and Theory of Generalization:

SUMMARY

$$E_{out} \simeq 0 \Rightarrow \underbrace{(E_{in} \simeq 0)}_{\text{learning algorithms}} + \underbrace{(E_{in} \simeq E_{out})}_{\text{theory of generalization}}$$

→ $$\text{Prob}\left[\,|E_{in}(g) - E_{out}(g)| > \epsilon\,\right] \leq \boxed{4 \cdot m_{\mathcal{H}}(2N) \cdot e^{-\frac{1}{8}\epsilon^2 N}}$$

δ

where, Growth function $m_{\mathcal{H}}(N) \leq \sum_{i=0}^{k-1} \binom{N}{i}$ with break point k

$$\boxed{m_{\mathcal{H}}(N) = O(N^{d_{vc}})} \Rightarrow m_{\mathcal{H}}(N) \leq \sum_{i=0}^{d_{vc}} \binom{N}{i} \quad \text{as} \quad d_{vc} = k-1$$

Ex: d-dimensional Perceptron ($d_{vc} = d+1$)

→ Generalization Bound: $\boxed{\text{PAC}}$

With probability / confidence $\geq 1-\delta$
we find, $E_{out} - E_{in} \leq \epsilon$

$$\epsilon = \sqrt{\frac{8}{N} \ln\left(\frac{4m_{\mathcal{H}}(2N)}{\delta}\right)}$$

$$\hookrightarrow E_{out} \leq E_{in} + \underbrace{\Omega(N, \mathcal{H}, \delta)}$$

← Generalization Bound

$N \geq 10\, d_{vc}$

practical exp

Approximation vs. Generalization:

$f: X \to Y$

$\langle x_1, y_1\rangle \cdots \langle x_N, y_N\rangle$

w/ TE

→ Hypothesis Set $\mathcal{H} = \{h_1, h_2 - - - - - - \}$

Learning Algo: $g \in \mathcal{H} \rightsquigarrow g \approx f$

- **Case-I**: more general $\mathcal{H} \Rightarrow$ better chances of approx. f.
- **Case-II**: less general $\mathcal{H} \Rightarrow$ better chance of gen.

$$E_{in}(g) \approx E_{out}(g)$$

best

Ideal $\Rightarrow \mathcal{H} = \{f\}$

$\bar{g}(x)$

→ How well $\boxed{\mathcal{H}}$ approximates f?

→ How can we find the best among $\mathcal{H}$?

$g(x) \approx \bar{g}(x)$

$\boxed{\text{bias}}$ vs $\boxed{\text{variance}}$

→ [TODAY'S LECTURE]

$$\underbrace{E_{out}(g^{(\mathcal{D})})} = \mathbb{E}_x[g^{(\mathcal{D})}(x) - f(x)]^2$$

$$\boxed{\mathbb{E}_x(f(x)) = \int_l^L f(x)\,dx} \quad \mathcal{D}' \rightarrow g'$$

$$\mathbb{E}_{\mathcal{D}}[E_{out}(g^{\mathcal{D}})] = \underbrace{\mathbb{E}_{\mathcal{D}}\Big[\mathbb{E}_x}\big(g^{\mathcal{D}}(x) - f(x)\big)^2\Big]$$

$$= \mathbb{E}_x\Big[\underbrace{\mathbb{E}_{\mathcal{D}}\big(g^{\mathcal{D}}(x) - f(x)\big)^2}\Big]$$

▷ Avg. Hypothesis

$$\mathcal{H} = \{h_1, \ldots, h_M\}$$

$$\hookrightarrow \bar{g}(x) = \underbrace{\mathbb{E}_{\mathcal{D}}[g^{\mathcal{D}}(x)]}$$

$$\boxed{\begin{array}{l}\text{Imagine } \mathcal{D}_1, \ldots, \mathcal{D}_K \\ \bar{g}(x) \simeq \dfrac{1}{K}\displaystyle\sum_{k=1}^{K} g^{\mathcal{D}_k}(x)\end{array}}$$

$$\mathbb{E}_{\mathcal{D}}\big[(g^{\mathcal{D}}(x) - f(x))^2\big]$$

$$= \mathbb{E}_{\mathcal{D}}\big[(\underbrace{g^{\mathcal{D}}(x) - \bar{g}(x)} + \underbrace{\bar{g}(x) - f(x)})^2\big]$$

$$= \mathbb{E}_{D}\big[(g^{\mathcal{D}}(x) - \bar{g}(x))^2 + (\bar{g}(x) - f(x))^2$$

$$\cancel{+ 2(g^{\mathcal{D}}(x) - \bar{g}(x)) \cdot \boxed{(\bar{g}(x) - f(x))}\big]}$$

$$\cancel{\Big[\underbrace{\mathbb{E}_{\mathcal{D}}[g^{\mathcal{D}}(x)} - \underline{\bar{g}(x)}]\Big] 2 \cdot (\bar{g}(x) - f(x))}$$

$$\not\ni \mathcal{D}$$

$$\cancel{\bar{g}(x) - \bar{g}(x)}$$

$$\mathbb{E}_{\mathcal{D}}\left[\left(g^{\mathcal{D}}(x)-f(x)\right)^2\right] = \underbrace{\mathbb{E}_{\mathcal{D}}\left[\left(g^{\mathcal{D}}(x)-\bar{g}(x)\right)^2\right]}_{\mathrm{var}(x)} + \underbrace{\left(\bar{g}(x)-f(x)\right)^2}_{\mathrm{bias}(x)}$$

$$\mathbb{E}_{\mathcal{D}}\left[\mathbb{E}_{x}\left[E_{out}(g^{\mathcal{D}})\right]\right]$$

$$= \mathbb{E}_{x}\left[\mathbb{E}_{\mathcal{D}}\left(g^{\mathcal{D}}(x)-\bar{g}(x)\right)^2\right] + \mathbb{E}_{x}\left[\left(\bar{g}(x)-f(x)\right)^2\right]$$

$$= \mathbb{E}_{x}\left[\mathrm{var}(x)\right] + \mathbb{E}_{x}\left(\mathrm{bias}(x)\right)$$

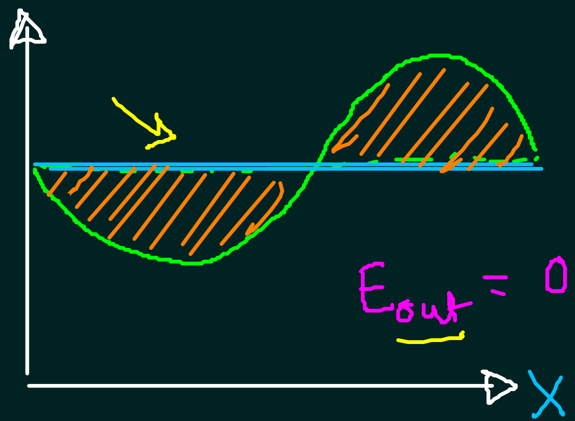
$$= \underbrace{\mathrm{var}}\uparrow + \underbrace{\mathrm{bias}}\downarrow \qquad \mathcal{H}\uparrow$$

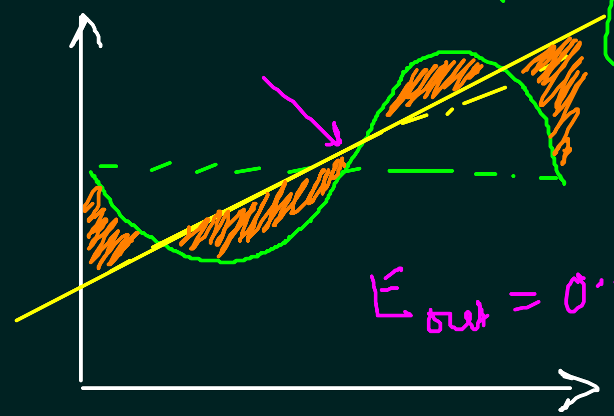
Example: $f : [-1, 1] \to \mathbb{R}^{[-1,-1]}$ where $f(x) = \sin(\pi x)$

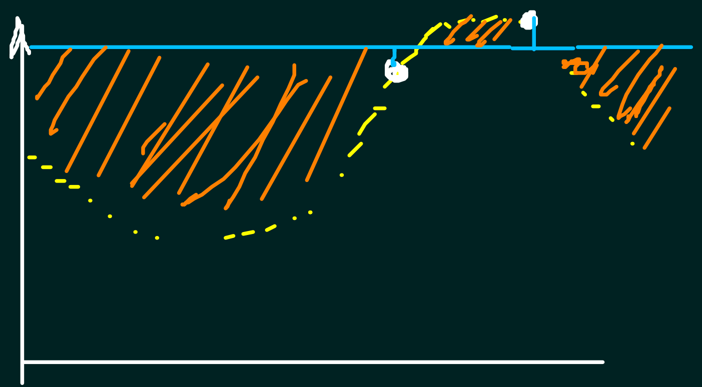
↳ $\#TE = 2 \leftarrow N$

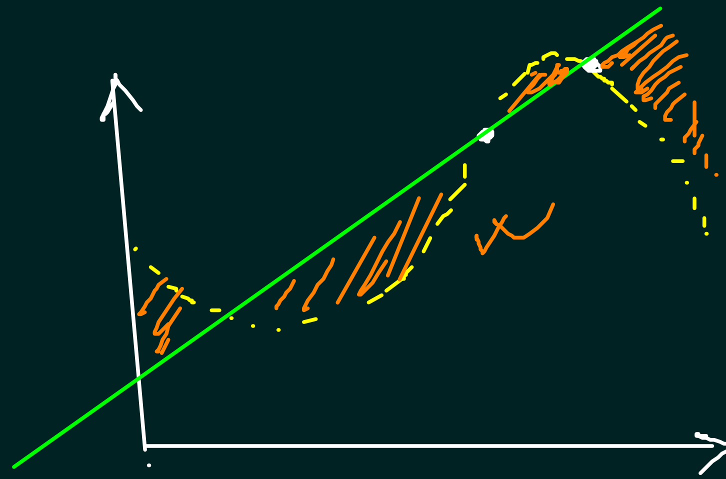
hyp: $\left\{ \begin{array}{l} \mathcal{H}_0 : \; h(x) = b \\ \mathcal{H}_1 : \; h(x) = ax + b \end{array} \right.$

$\boxed{(2)}$





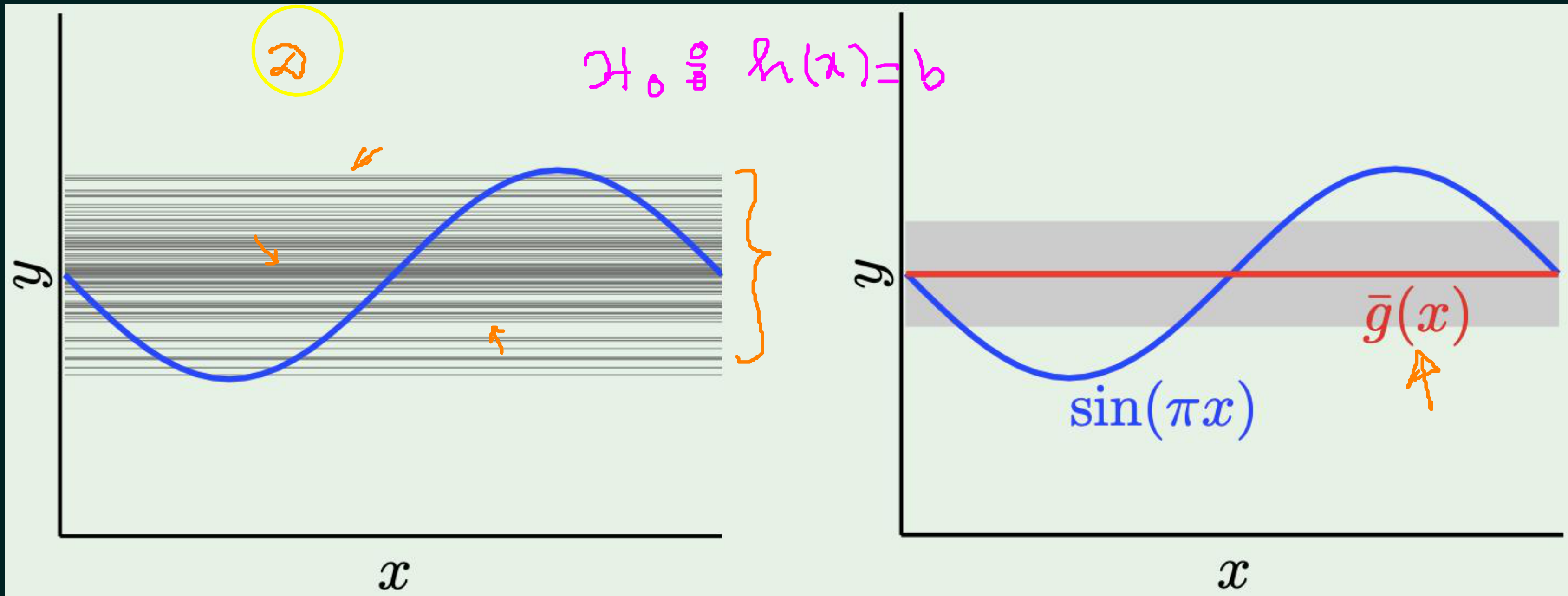




$\left.\begin{array}{l} \mathcal{H} \uparrow \\ \text{bias} \downarrow \\ \text{var} \uparrow \end{array}\right\}$

(Small Model (complex))

$E_{out} \approx 0 \longrightarrow \boxed{\mathcal{H}_0}$ vs $\mathcal{H}_1$ $(N = 2)$



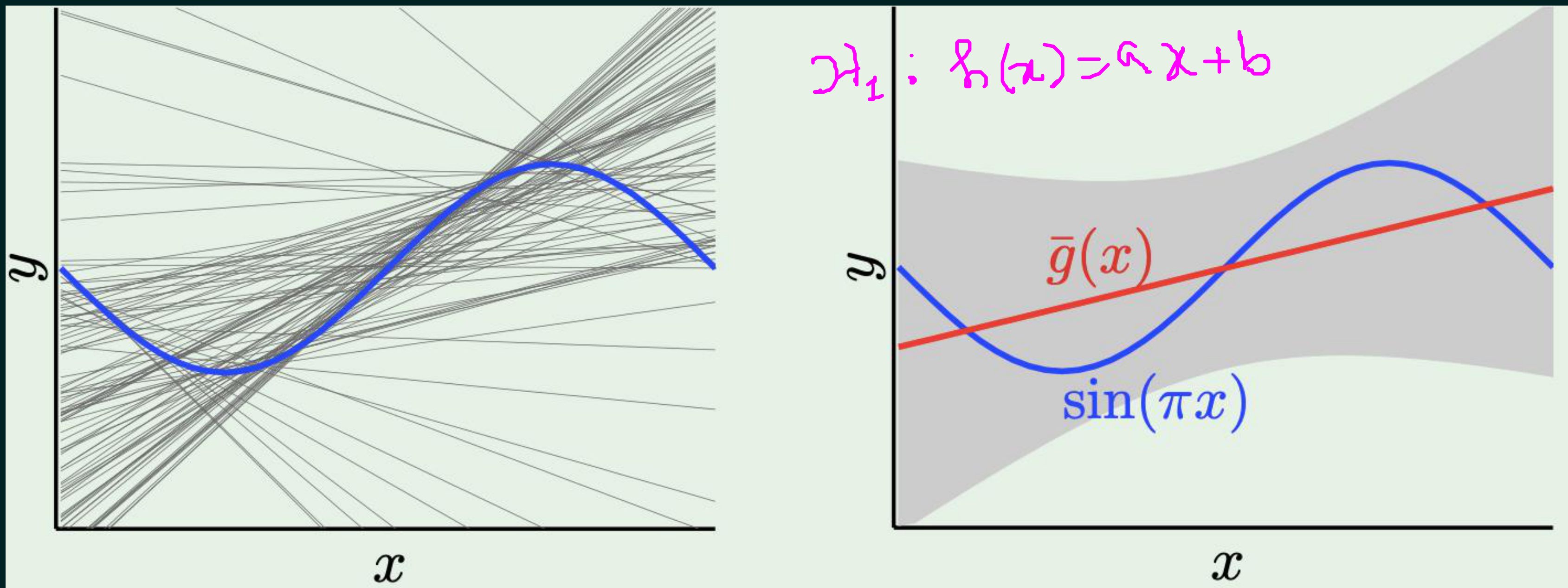
var

bias = 0.50

Var = 0.25

0.75

f



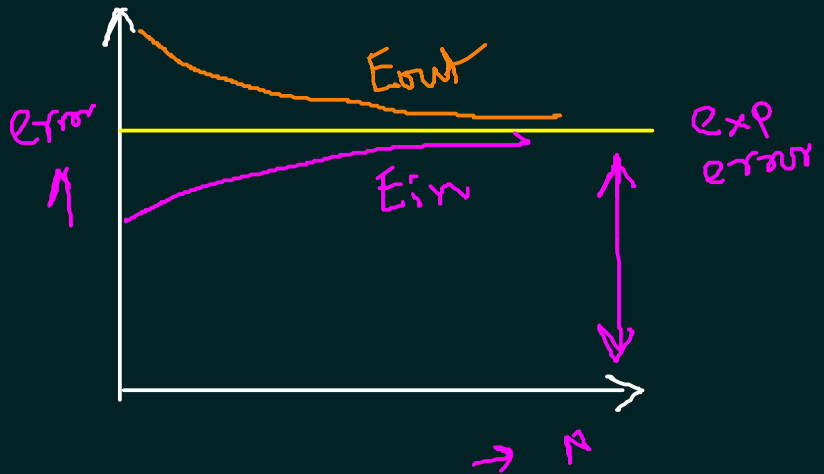
var

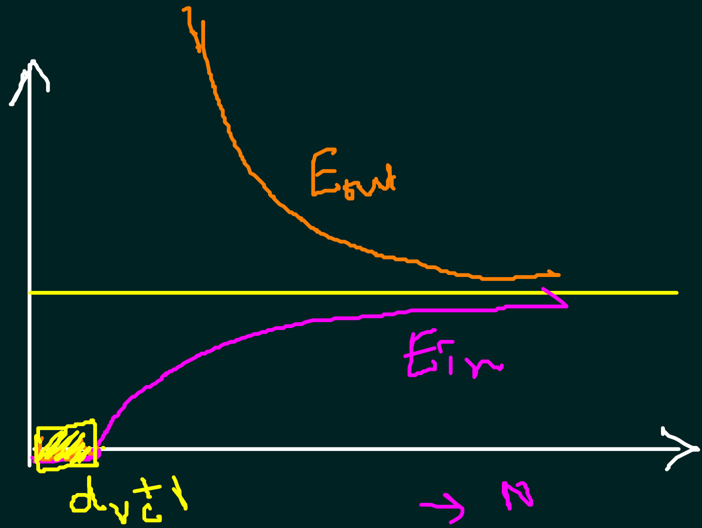
bias = 0.21

var = 1.69

1.90

$E_x\left[\bar{g}(x) - f(x)\right]^2$





sq-error

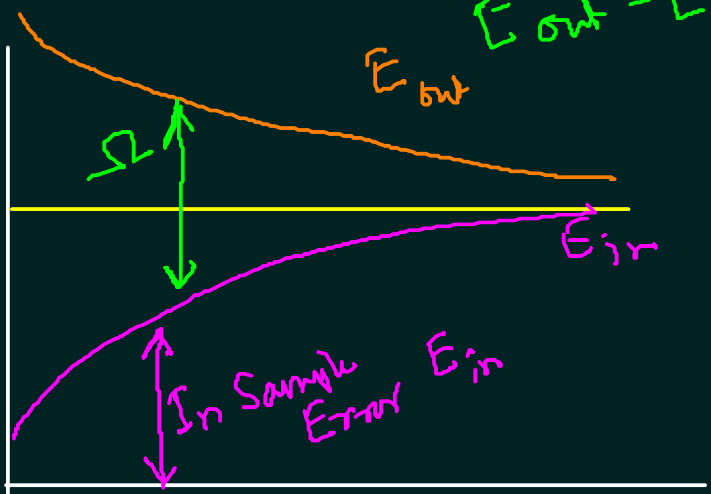
$\mathcal{H}_0$ Simple Model

$\mathcal{H}_1$ Complex Model

$\text{bias}(x)$

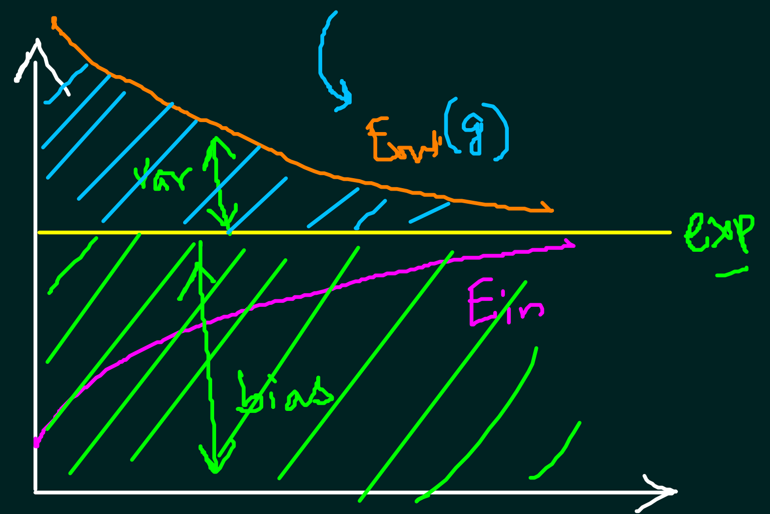
$= [\bar{g}(x) - f(x)]^2$

$E_{out} - E_{in} \le \Omega$



Ex: Linear Reg.

σ^2



$$E_{out}(g) \le E_{in}(g) + \Omega$$

$$E_{out}(g) \le \text{bias} + \text{var}$$