Recap:

↳ classification: H complexity $\equiv$ $d_{VC}(H)$

Regression : " " " $\Rightarrow$ Bias variance tradeoff

↳ $\mathcal{D} = \{(\underline{x}_1, y_1) \cdots (\underline{x}_n, y_n)\}$

$$\boxed{\begin{aligned} &\text{Output hypo: } g^D : \hat{y} = g^D(\underline{x}) \\ &\text{Avg hypo: } \bar{g}(\underline{x}) = \mathbb{E}_D[g^D(\underline{x})] \\ &\text{bias}(\underline{x}) = (\bar{g}(\underline{x}) - f(\underline{x}))^2 \\ &\text{var}(\underline{x}) = \mathbb{E}[(g^D(\underline{x}) - \bar{g}(\underline{x}))^2] \end{aligned}}$$

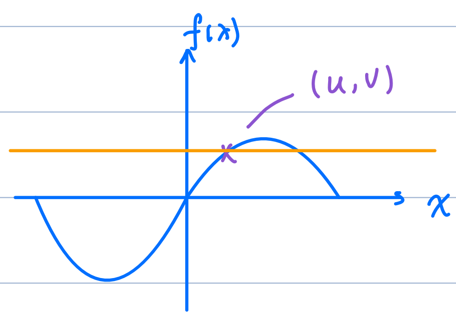
e.g. $d = 1$, $x \in [-1, 1]$

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

H : constant func

$\mathcal{D} = \{(u, v)\}$ // only one data , $N = 1$

and $u \sim \mathcal{U}(-1, 1)$ // u is randomly chose accord to $\mathcal{U}(-1, 1)$



the out. hypo.

$g^D(x) = v$, $\forall x$

$\bar{g}(x) = \mathbb{E}_u[g^D(x)] = \mathbb{E}_u[v] = \mathbb{E}_u[\sin(\pi, u)] = \int_{-1}^{1} \frac{1}{2} \sin(\pi u)\, du = 0$

$\therefore$ bias$(x) = (0 - \sin(\pi x))^2 = \sin^2(\pi x)$

var$(x) = \mathbb{E}_u[(v - 0)^2] = \int_{-1}^{1} \frac{1}{2} \sin^2(\pi u)\, du = \frac{1}{2}$

↳ Avg test error for sample $\underline{x}$: $\Delta(\underline{x})$

$$\Delta(\underline{x}) \triangleq \mathbb{E}_D\left[\left(g^D(\underline{x}) - f(\underline{x}) \right)^2 \right]$$

output truth grd truth

e.g. In the prev example:

$$\Delta(x) = \mathbb{E}_u\left[(v - \sin(\pi x))^2 \right]$$

$$= \mathbb{E}_u\left[v^2 + \sin^2(\pi x) - 2v\sin(\pi x) \right]$$

$$= \mathbb{E}[v^2] + \sin^2(\pi x) - 2\mathbb{E}[v]\sin(\pi x)$$

$$= \frac{1}{2} + \sin^2(\pi x) - 0$$

$$= \text{bias}(x) + \text{var}(x)$$

$\therefore$ In general. $\Delta(x) = \text{bias}(x) + \text{var}(x)$

prf: $\Delta(\underline{x}) \triangleq \mathbb{E}_D\left[\left(g^D(\underline{x}) - f(\underline{x}) \right)^2 \right]$

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

$$= \mathbb{E}_D\Big[\left(\bar{g}(\underline{x}) - f(\underline{x})\right)^2 + \left(g^D(\underline{x}) - \bar{g}(\underline{x})\right)^2$$

$$+ 2\left(\bar{g}(\underline{x}) - f(\underline{x})\right)\cancel{\left(g^D(\underline{x}) - \bar{g}(\underline{x})\right)} \Big] \rightarrow 0$$

$$= \text{bias}(\underline{x}) + \text{var}(\underline{x})$$

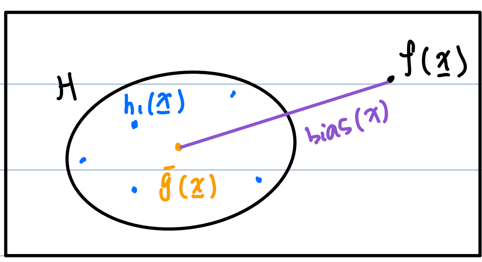
2. Bias-variance trade off for some given $\underline{x}$:

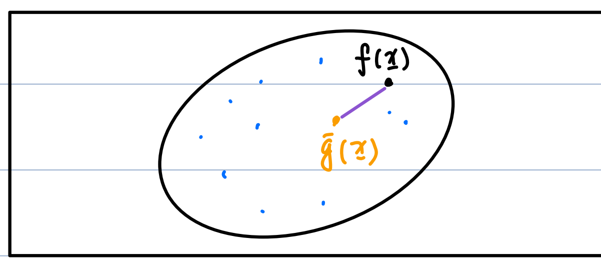
↳ Bias: How well can H approx. the target function $f(\underline{x})$

Variance: How much does the output hypo. $g^D(\underline{x})$ change with $\mathcal{D}$.

For some $\underline{x}$

space of labels





small H. $\Rightarrow$ { bias larger

large H $\Rightarrow$ { bias smaller

| var smaller | var larger

↳ <u>Averaging over $\underline{x}$</u>:

- $E_{out}(g^D) = \mathbb{E}_{\underline{x}}[(g^D(\underline{x}) - f(\underline{x}))^2]$, $\underline{x} \sim P(\underline{x})$

Note: $E_{out}(g^D)$ is random (depends on $\mathcal{D}$)

- $\overline{E_{out}} \triangleq \mathbb{E}_D[E_{out}(g^D)]$

$$= \mathbb{E}_D[\mathbb{E}_{\underline{x}}[(g^D(\underline{x}) - f(\underline{x}))^2]]$$

where $\underline{x}$ and $\mathcal{D}$ are independent.

$$= \mathbb{E}_{\underline{x}}[\mathbb{E}_D[(g^D(\underline{x}) - f(\underline{x}))^2]]$$

$$= \mathbb{E}_{\underline{x}}[\Delta(\underline{x})]$$

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

$$\triangleq \overline{\text{bias}} + \overline{\text{var}}$$

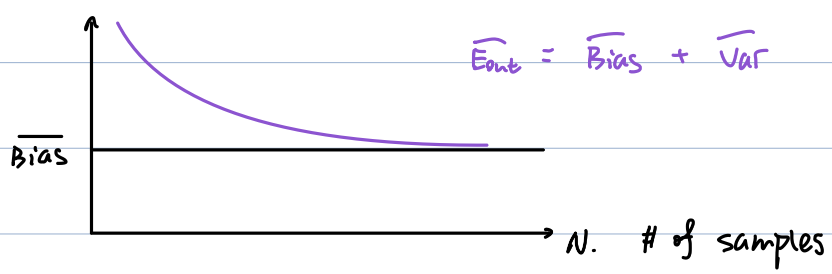
$$\overline{E_{out}} = \overline{\text{bias}} + \overline{\text{var}}$$

e.g. In prev example:

$\overline{\text{bias}} = \mathbb{E}_x[\underbrace{\sin^2(\pi x)}_{\text{bias}(x)}] = \frac{1}{2}$, $\overline{\text{Var}} = \mathbb{E}_x[\frac{1}{2}] = \frac{1}{2}$

$\therefore \overline{E_{out}} = \frac{1}{2} + \frac{1}{2} = 1$

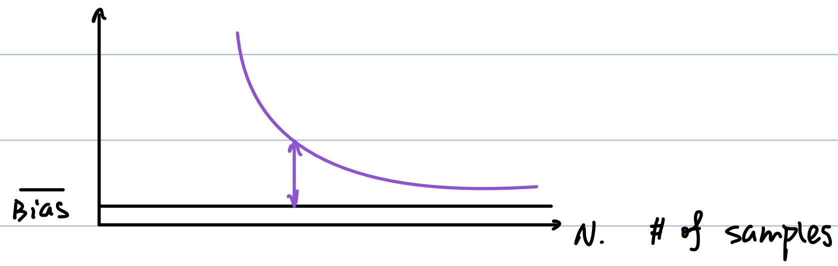
- In general:



(LLNs)

$$g^D(\underline{x}) \to \mathbb{E}_D[g^D(\underline{x})]$$

· More complex $\mathcal{H}$:



※ If has more datapoints, can have more complex model

e.g. (prev) but now $N = 2$.