Tptr6r

Lecture-13 Tptr6r

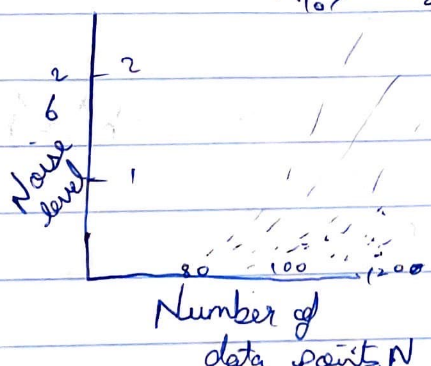
- Overfit Measure: $E_{out}(H_{10}) - E_{out}(H_2)$

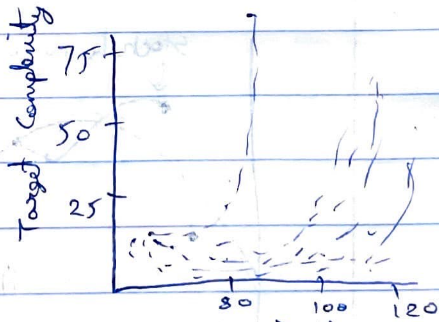
Learning Scenario Parameters:
- Noise level σ^2 (Variance)
- target Complexity
-

Favour H_{10} only if you have large dataset.

- If the noise increases, prefer simpler model.

$H_{10} > H_2$





$H_2 > H_{10}$.

Noise level σ^2 (Variance) ↑ ↑ Overfitting
Target compexity Q_f ↑ ↑ Overfitting:
Complex no. of data points ↑ ↓ Overfitting.

Noise: The Part of y we cannot model
stochiastic Noise.
- Error added to target function.

$y = f(x) + \varepsilon$

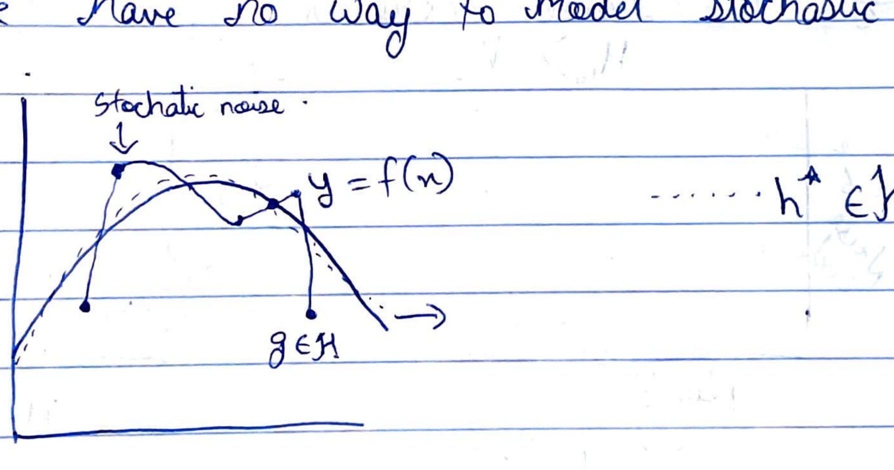
$f(x)$: o/p of target func^n

ε: noise stochiastic Noise

Even on test point

$\hat{y} = g(x)$.

★ We have no way to model stochastic noise.



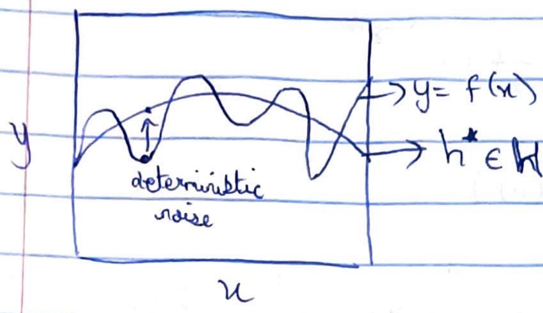
~~If~~ Instead of making the right curve (h^*) we end up making g ~~so~~ because of stochatic noise.

★ Stochastic & Deterministic Noise.

~~Stochastic~~ #

Deterministic Noise -

the other ~~point~~ part of y which cannot model.



We get this deterministic noise because we have already fixed the hypothesis set.
& we dont know the target func^n. ~~Bec~~

Choice of hypothesis set give arise to deterministic noise.

$$\begin{cases} y_n = f(x_n) \\ y_n = h^*(x_n) + \text{deterministic choice} \end{cases}$$

target complexity ↑ Data points ↓ then simpler model

Stochastic noise	Deterministic noise.
• random measurement errors	• H cannot model f.
• It changes if we measure y_n again	• Value remains const^n even if we measure y_n again.
• Independent of H	• Dependent of H

$\mathbb{E}$ = Expected value.

Bias Variance Decomposition with stochastic noise.

$$\mathbb{E}\left[E_{out}(x)\right] = bias(x) + variance(x)$$

Expected value.

$$bias(x) = \left(\bar{g}(x) - f(x)\right)^2$$

$$Var(x) = var(g(x))$$

$$E_{out}(x) = \left(g(x) - y\right)^2, \text{ where } y = f(x) + \varepsilon$$

$$= \left(g(x) - f(x) - \varepsilon\right)^2$$

$$= \left(g(x) - f(x)\right)^2 - 2\varepsilon\left(g(x) - f(x)\right) + \varepsilon^2$$

$$\mathbb{E}_x\left[E_{out}\right] = \mathbb{E}_x\left[\left(g(x) - f(x)\right)^2 - 2\varepsilon\left(g(x) - f(x)\right) + \varepsilon^2\right]$$

Expected Value.

$$* \ \mathbb{E}[\varepsilon] = 0 \Rightarrow$$

$$\mathbb{E}_x\left[E_{out}\right] = \mathbb{E}_x\left[\left(g(x) - f(x)\right)^2\right] + \mathbb{E}_x\left[\varepsilon^2\right]$$

$$= \mathbb{E}_x\left[bias(x) + Var(x)\right] + \mathbb{E}_x\left[\varepsilon^2\right]$$

$$E_{out} = bias + \varepsilon^2 + Var.$$

deterministic noise — stochastic noise. Indirect impact of noise

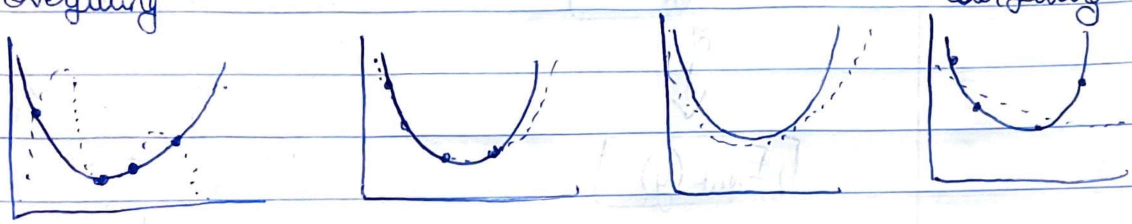
* We have to find the best value of λ that will give us lower value pair of bias & variance.

$$\nabla E_{aug}(w) = 0.$$

$$W_{reg} = (Z^T Z + \lambda I)^{-1} Z^T y.$$

Minimizing $E_{in}(w) + \frac{\lambda}{N} w^T w$

$\lambda = 0$ overfitting | $\lambda = 0.0001$ | $\lambda = 0.01$ | $\lambda = 1.$ Underfitting



More noise demands more Regularization.