hbQg8F

Lecture -14 hbQg8F

How to pick? Validation : Estimate.

Goal of Learning: Pick which minimizes $E_{out}(g)$

$E_{out}(g) = E_{in}(g) +$ model complexity penalty

Regularization :- Use λ to control the penalty for model complexity.
How to pick λ?

Estimate $E_{out}(g)$ 'directly'.

$E_{test}(g)$ is an unbiased estimator of $E_{out}(g)$.

$$\text{Variance of } [E_{test}(g)] = \text{Var}\left[\frac{1}{K}\sum_{k=1}^{K} e_k\right]$$

$$= \frac{1}{K^2}\text{Var}\left[\sum_{k=1}^{K} e_k\right]$$

$$= \frac{1}{K^2}\sum_{k=1}^{K}\text{Var}[e_k]$$

$$= \frac{1}{K^2}\sum_{k=1}^{K}\text{Var}[\text{error}(g(x), f(x))]$$

$$= \frac{1}{K}\text{Var}[\text{error}(g(x), f(x))]$$

- Using $E_{val}(g^-)$ to Estimate $E_{out}(g)$.

From VC analysis:

$$E_{out}(g) \leq E_{in}(g) + O\left(\sqrt{\frac{d_{vc}}{N} \log N}\right)$$

From Hoeffding Inequality:

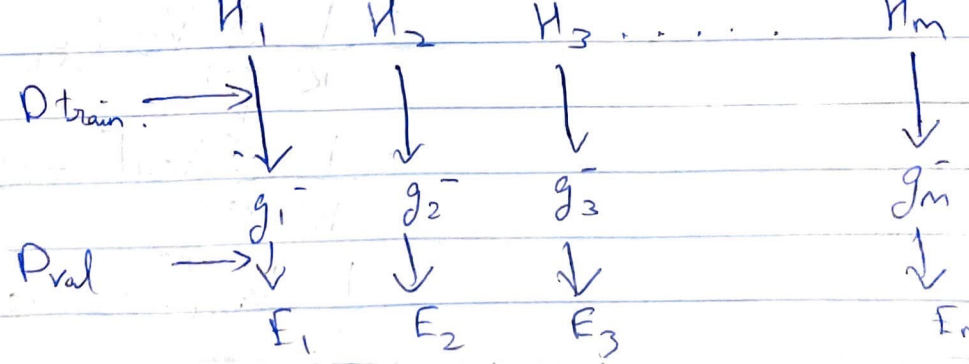
Typically, $E_{out}(g) \leq E_{out}(g^-) \leq E_{val}(g^-) + O\left(\frac{1}{\sqrt{K}}\right)$

small K ~~large K~~ large K

We need to find good value of K.

such that $E_{val}(g)$ should be a good estimate of $E_{out}(g^-)$ & $E_{out}(g)$ should be a good estimate of $E_{val}(g^-)$.

- Use validation to Evaluate Models.



pick the model with smallest $E_{val}(g_*^-)$

Validation error.

* Cross-Validation [or k = 1]:-

leave 1 out for cross-validation.

$D = \{(x_1, y_1) \ldots (x_n, y_n)\}$

When K = 1

N - 1 data points.

1st Iteration	2nd Iteration	n-1th iteration
(N-1) points	$x_1, \ldots x_n$	$(x_2, y_2) \ldots$
$E_{in} = x_1, x_2, \ldots x_{n-1}$		(x_n, y_n)
$E_{val} = x_n$	x_{n-1}	Holding 1st data point.
		$Val = (x_1, y_1)$

g_1^- g_2^- ..... g_n^-

(x_1, y_1) (x_2, y_2) (x_n, y_n)

e_1 e_2 e_n

average.

E_{CV}

(Cross Validation error)

$$E_{cv} = \frac{1}{N} \sum_{n=1}^{N} e_n.$$

It is a good estimate of the out of sample error when you train on N-1 data points The variance can be large in this case, Variance of these out of sample will be high.

$$\mathbb{E}_{D_n}[E_{cv}] = E_{out}[N-1]$$

But we can't implement this on large dataset of 10,000 so we take 4-5 random cross Validation.

D (Dataset).

$L \rightarrow K = \frac{N}{L}$

D (Random partition),

D_1 D_2 D_L

L for (cross Validation) → D D.

→ Dval D_1 D_2 .... D_L.