CS19BTECH11020

_/_/__

Assignment - 4

Q1) (a) Given (x_n, t_n), $n = 1, \dots N$

Error f^n $E_D(w) = \frac{1}{2}\sum_{n=1}^{N} g_n (t_n - w^T\phi(x_n))^2$

To minimize error, derivative should be equal to zero.

$$\frac{\partial}{\partial w} E_D(w) = 0$$

$$\frac{\partial}{\partial w} \frac{1}{2} \Sigma g_n (t_n - w^T\phi(x_n))^2$$

$$= -\sum_{n=1}^{N} g_n (t_n - w^T\phi(x_n)) \cdot \phi(x_n) = 0$$

$$\Rightarrow -\sum_{n=1}^{N} g_n \phi(x_n) t_n + \Sigma g_n w^T \phi(x_n) \phi(x_n) = 0$$

$$\sum_{n=1}^{N} g_n \phi(x_n) \phi(x_n)^T) w = \Sigma g_n \phi(x_n) t_n$$

$$W = \left(\sum_{n=1}^{N} g_n \phi(x_n) \phi(x_n)^T\right)^{-1} \left(\sum_{n=1}^{N} g_n \phi(x_n) t_n\right)$$

(b) For data dependant noise varience, let $t_i = w^T x_i + \epsilon_i$

here $\epsilon_i = \mathcal{N}(0, \sigma_i^2)$

$$t_i = w^T x_i + \mathcal{N}(0, \sigma_i^2)$$
$$= \mathcal{N}(w^T x_i, \sigma_i^2)$$

From maximum likelihood

$$\text{maxlike}(w/\text{data}) = \arg\max_w \prod_{i=1}^{N} P(t_i/w)$$

$$= \arg\max_{w} \prod_{i=1} \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{\left(-\frac{(t_i - w^T x_i)^2}{2\sigma_i^2}\right)}$$

Log Likelihood is max for

$$w = \arg\max_{w} \sum_{i=1} N \log \frac{1}{\sqrt{2\pi\sigma_i^2}} + \left(-\sum_{i=1}^{N} \frac{(t_i - w^T x_i)^2}{2\sigma_i^2}\right)$$

$$= \arg\max_{w} \sum_{i=1}^{N} \frac{(t_i - w^T x_i)^2}{2\sigma_i^2}$$

Comparing abov eq with

$$E_D(w) = \arg\min_{w} \frac{1}{2} \sum g_i (t_i - w^T x_i)^2$$

$$\therefore \boxed{g_i = \frac{1}{\sigma_i^2}}$$

$\therefore$ for data depandant noise varience

$$E_D(w) = \frac{1}{2} \sum \frac{1}{\sigma_i^2} (t_i - w^T \phi(x_i))^2$$

(ii) In case of replicated data points, Let i^{th} data point repeats g_i times then

$$E_D(w) = \sum_{i=1}^{N} g_i \frac{(t_i - w^T x_i)^2}{2} = \frac{1}{2} \sum_{i=1}^{N} g_i (t_i - w^T \phi(x_i))^2$$

$\therefore$ we get same error f^n g

$$E_D(w) = \frac{1}{2} \sum_{i=1}^{N} g_i (t_i - w^T \phi(x_i))^2 \qquad \text{where } g_i > 0$$

replication has same effect as giving weights to datapoints.

② Bayes optimal estimate:

$$h_{Bayes} = \text{argmax}\ (h_{Bayes}(F), (L), R))$$

$$\sum_{h_i \in H} P(F/h_i)P(h_i/D) = 0.4 + 0 + 0 + 0 + 0 = 0.4$$

$$\sum_{h_i \in H} P(L/H_i)P(h_i/D) = 0.2 + 0 + 0.1 + 0.2 = 0.5 \quad \leftarrow \text{max}$$

$$\sum_{h_i \in H} P(R/h_i)P(h_i/D) = 0 + 0 + 0.1 + 0 + 0 = 0.1$$

⇒ The robot will turn $\boxed{\text{Left}}$ (L)

Map estimate

$$h_{map} = \text{argmax}\ P(h_i/D)$$

$$= \max\{0.4, 0.2, 0.1, 0.1, 0.2\}$$
$$= 0.4$$

when $P(h_i/D) = 0.4$, $P(F/h_i) = 1$

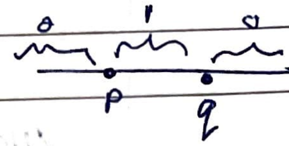
→ The robot will be moving $\boxed{\text{forward}}$ (F)

∴ The map estimate and Bayes optimal estimate are not same.

③ data $\in R^1$

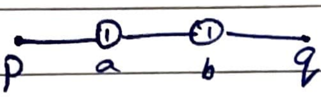
H is parameterized by $\{P, q\}$

x is classified 1 iff $P < x < q$ →

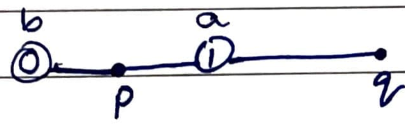


All the possible labellings of 2 points a,b are

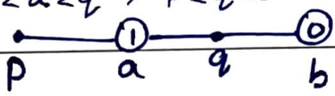
(i) $P < a < q$, $p < b < q$



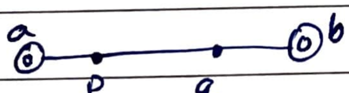
(ii) $P < a < q$, $b < p < q$



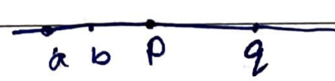
(iii) $p < a < q$, $p < q < b$



(iv) $a < p < q$, $p < q < b$



(v) $a < b < p < q$



∴ For all two points configurations we can represent on straight line and we can distinguish by hypothesis.

If we consider 3 points 1 — 0 — 1, They can't be classified by given hypothesis space H.

∴ VC dimension of H is $\boxed{2}$

(4) $x = [x_1, x_2, \ldots x_D)$

$$y(x, w) = w_0 + \sum_{k=1}^{D} w_k x_k$$

Sum of squared error $\Rightarrow$ $E(w) = \frac{1}{2} \sum_{i=1}^{N} (y(x_i, w) - t_i)^2$

Gaussian noise ϵ_k is added to each x_k

$$\therefore y'(x, w) = w_0 + \sum_{k=1}^{D} w_k (x_k + \epsilon_k)$$

$$= w_0 + \sum_{k=1}^{D} w_k x_k + \sum_{k=1}^{D} w_k \epsilon_k$$

$$y'(x, w) = y(x, w) + \sum_{k=1}^{D} w_k \epsilon_k$$

$\therefore$ New error f^n

$$E'(w) = \frac{1}{2} \sum_{i=1}^{N} (y'(x_i, w) - t_i)^2$$

$$= \frac{1}{2} \sum_{i=1}^{N} \left(y(x_i, w) + \sum_{k=1}^{D} w_k \epsilon_k - t_i \right)^2$$

$$E'(w) = \frac{1}{2} \sum_{i=1}^{N} \left\{ y(x_i, w) - t_i)^2 + \left(\sum w_k \epsilon_k\right)^2 - 2\left(\sum w_k \epsilon_k\right)(y(x_i, w) - t_i) \right.$$

Expectation on both sides

$$E[E'(w)] = \frac{1}{2} \sum_{i=1}^{N} \left\{ (y(x_i, w) - t_i)^2 + E\left[\left(\sum_{k=1}^{D} w_k \epsilon_k\right)^2 \right] - 2(y(x_i, w) - t_i) \sum_{k=1}^{D} w_k E[(\epsilon_k)] \right.$$

$$= \frac{1}{2} \sum_{i=1}^{N} \left\{ (y(x_i, w) - t_i)^2 + \left[\sum_{k=1}^{D} \sum_{k'=1}^{D} w_k w_{k'} E[\epsilon_k \epsilon_{k'}] \right] \right\}$$

$$= \frac{1}{2} \sum_{i=1}^{N} \left[(y(x_i, w) - t_i)^2 + \sum_{k=1}^{D} w_k^2 \right]$$

$$E[E'(w)] = E(w) + N/2 \sum_{k=1}^{D} w_k^2$$

$\therefore$ The relation between sum of squares averaged over noisy data & standard sum of squared error is

$$\boxed{E_D'(w) = E_D(w) + \frac{N}{2} \sum_{k=1}^{n} w_k^2}$$