FOML

SRI HARI. M

CS19BTECH11039

1. Given error function,

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

a. Expression to minimise the error function,

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

$$\Rightarrow \frac{\partial}{\partial w}\left(\frac{1}{2} \sum_{n=1}^{N} g_n\left(t_n - w^T \phi(x_n)\right)^2\right) = 0$$

$$\Rightarrow 2 \cdot \frac{1}{2} \cdot (-1) \cdot \Sigma g_n\left(t_n - w^T \phi(x_n)\right) \cdot \phi(x_n) = 0$$

$$\Rightarrow -\Sigma g_n\left(t_n - w^T \phi(x_n)\right) \cdot \phi(x_n) = 0$$

$$\Rightarrow \Sigma g_n t_n \phi(x_n) = \left(\Sigma g_n \phi(x_n) \phi(x_n)^T\right) w$$

$$\Rightarrow w = \left(\Sigma g_n t_n \phi(x_n)\right)\left(\Sigma g_n \phi(x_n) \phi(x_n)^T\right)^{-1}$$

b. Guassian noise is added independently to each of the input variables x_k

$$\Rightarrow t_i \sim N(\bar{w}x_i, \sigma_i^2)$$

$$\text{argmax } \mathcal{L}(\theta/x) = \text{argmax } \prod P_{(\mu,\sigma^2, x)}$$

Maximising the log likelihood,

$$\text{argmax } \log(\mathcal{L}(\theta/x)) = \text{argmax } \log(\prod P_{(\mu,\sigma^2, x)})$$

$$= \text{argmax } \sum \left[-\frac{1}{2}\log(2\pi\sigma^2) - \frac{(x-\mu)^2}{2\sigma^2} \right]$$

$$= \text{argmax } g(\mu, \sigma^2)$$

Now we need to maximise, μ and σ^2

W.K.T $\quad \mu = \frac{\sum x}{n}$

$$\sigma^2 = \frac{\sum (x-\mu)^2}{n}$$

$$\Rightarrow \text{argmax} = \sum \frac{(\bar{w}x - t)^2}{2\sigma^2}$$

from a, $\frac{1}{\sigma^2} = g^{-1}$

Setting this value, we get the original expression.

For replicated data points, if we add a row, the expression looks like

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

which is similar to given $E_D(w)$

FOML

SRI HARI. M
CS19BTECH11039

2, Given 5 hypothesis,

Bayes Optimal Estimate makes most probable predictions using given Space of hypothesis (h)

$$\underset{y}{\text{Argmax}} \text{ of } \sum P(y/h)\,P(h/D)$$

$\sum P(F/h)P(h/D) = 0.4 \times 1 + 0$

$\sum P(L/h)P(h/D) = 0.2 \times 1 + 0.1 \times 1 + 0.2 \times 1$

$\sum P(R/h)P(h/D) = 0.1 \times 1$

Max of the three is left; So according to Bayes optimal estimate, robot goes left.

Map Estimate chooses most probable hypothesis and then evaluates the data.

$$\underset{y}{\text{Argmax}} \text{ of } P(y/h)P(h/D)$$

$= \max(0.4, 0.2, 0.1, 0.2, 0.1)$

$= 0.4$

So, according to map estimate, robot goes forward.

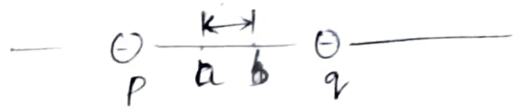
They aren't Same. Hence justified.

3. VC Dimention:

Considered a data setup of one-dimensional data where hypothesis space is parametrized by p and q

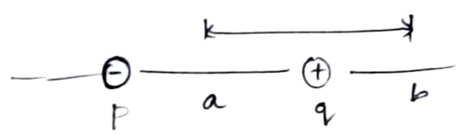
$x \to 1 \iff p < x < q$

R.T.F VC dimension of hypothesis.

① 

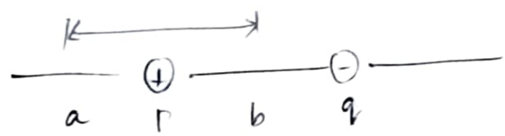
choose a, b such that $p < a < b < q$

$\iff$ So p, q are −

② 

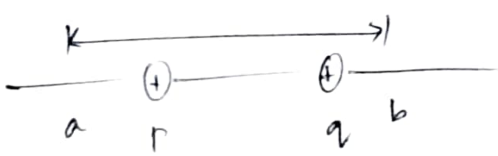
$\Rightarrow$ $p \to -ve$ $q \to +ve$

Choose a, b such that $p < a < q < b$

③ 

$\Rightarrow$ $p \to +ve$, $q \to -ve$

Choose a, b Such that $a < p < b < q$

④ 

$\Rightarrow$ $p \to +ve$, $q \to +ve$

choose a, b Such that $a < p < q < b$.

We clearly demonstrated using two dimension. But with 3 dimensions it is not possible, with configuration − + − / + − + . $\therefore$ VC dimension of hypothesis = 2.

(4) Given D-dimensional data

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

Supposed that, Guassian noise is added independently to each of the input variables.

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

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

let initial sum of squares error be

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

E' be sum of squares errors averaged over the noisy data.

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

$$E'(w) = \frac{1}{2} \sum_{i=1}^{N} (y(x,w) + \sum w_k \varepsilon_k - t_i)^2$$

$$= \frac{1}{2} \sum_{i=1}^{N} ((y(x,w) - t_i) + \sum w_k \varepsilon_k)^2$$

$$= E(w) + \frac{1}{2} \sum_{i=1}^{N} (\sum w_k \varepsilon_k)^2 + \sum_{i=1}^{N} (y(x,w) - t_i)(\sum w_k \varepsilon_k)$$

Given mean <u>noise</u> and variance of Guassian noise as 0,

$$= E(w) + \sum_{i=1}^{N} (y(x,w) - t_i)(0) + \frac{N\sigma^2}{2} \sum w_k^2$$

$$= E(w) + \frac{N(0)}{2} \sum w_k^2$$

$$E'(w) = E(w)$$

So, the sum of squares error are same before and after adding the guassian noise.