ECE 792 (061) HW2 SOMSHUBHRA ROY

1) R.T.P

a) · maximising likelihood function

$P(Y_D \mid x, w, \alpha)$

is equivalent to minimising

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

for a multi output neural network (Regression model)

where α is a precision parameter.

PROOF

given enough data

$Y_D \sim N(Y(x, w), \alpha^{-1} I)$

Here Y_D, $Y(x, w)$, x, w are vectors of N dimension corresponding to N training samples.

where $y(x, w)$ ie the predicted y is the mean of the distribution around which Y_D varies with a variance of α^{-1} for each sample.

$$\therefore P(Y_D \mid x, w, \alpha) = N(Y_D \mid Y(x, w), \alpha^{-1} I)$$

$$= \prod_{i=1}^{N} \frac{1}{\sqrt{2(\alpha^{-1} I)^2 \pi}} e^{-\frac{(Y_{D_i} - Y(x_i - w))^2}{2(\alpha^{-1} I)^2}}$$

(considering N independent training samples)

$$\therefore \text{R.T.F } w_{ML} = \underset{w}{\text{argmax}} \left(\prod_{i=1}^{N} \frac{1}{\sqrt{2(\alpha^{-1} I)^2 \pi}} e^{-\frac{(Y_{D_i} - y(x_i - w))^2}{2(\alpha^{-1} I)^2}} \right)$$

$$\therefore \ln(P(Y_D \mid x, w, \alpha)) = -\frac{\alpha}{2} \sum_{i=1}^{N} \| Y(x_i, w) - Y_{D_i} \|^2 + \frac{N}{2} \ln(\alpha) - \frac{N}{2} \ln(2\pi)$$

$\because \sigma \in R$ and is non negative, $\sum_{i=1}^{N} \| Y(x_i, w) - Y_{D_i} \|^2$ is also a non negative Real number.

$\because \alpha \in R$ and is non negative

∴ maximising P is equivalent to maximising ln(P)
which is equivalent to minimising the only negative w dependent term.

ie minimising $\frac{1}{2}||y(x_i, w) - y_{D_i}||^2$ [∵ $\beta \in R^+$]

[Proved]

b) maximising likelihood in terms of α

$$\frac{1}{\alpha_{ML}} = \frac{1}{N}\sum_{n=1}^{N} ||y_{D_i} - y(x_i, w_{ML})||^2$$

w_{ML} was found by maximising P( ) on w

---

2) for a multiclass classification

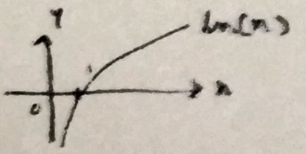
$$P(y_{D_k} = 1 | x) = y_k(x_i, w) = \frac{e^{w_k^T x}}{\sum_j e^{w_j^T x}}$$ (likelihood function) (softmax)

$$P(y_{D_k} = 1 | w_1, w_2 \cdots w_K) = \prod_{i=1}^{N}\prod_{k=1}^{K} y_k(x_i, w)^{y_{D_k}}$$

$$E(w_1, w_2 \cdots w_K) = -\ln p(y_{D_k} = 1 | w_1 \cdots w_K) = -\sum_{i=1}^{N}\sum_{k=1}^{K} y_{D_k} \ln(y_k(x_i,$$

$$\therefore E(w) = -\ln p(y_{D_k} | w) = -\sum_{i=1}^{N}\sum_{k=1}^{K} y_{D_k} \ln(y_k(x_i, w))$$

∵ ln( ) is monotonically increasing..



∴ minimising E(w) is equivalent to maximising likelihood (f