Linear Regression: for 17/02/24

A line with gaussian noise around it $\leftarrow y_i = \underline{w^T x_i} + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2)$ → ①

$\swarrow Y_i \sim N(w^T x_i, \sigma^2)$ → ②

$\Rightarrow$ ① & ② are Same.

Each point is from a gaussian distribution with mean $w^T x_i$ and Variance σ^2.

<u>Assumptions</u>: ① I/P & O/P are linearly related.
② Noise follows gaussian distribution.

$$P(y_1 x_1 \ldots y_n x_n / w) = \prod_{i=1} P(y_i / x_i, w)$$

$$P(y_i / \vec{x_i}, w) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(w^T x_i - y_i)^2}{2\sigma^2}}$$ → Distribution of y given x & w.

Two approaches/ways of finding w
① MLE ② MAP

<u>MLE</u>:

$$w = \underset{w}{\operatorname{argmax}} \prod_{i=1}^{n} P(y_i / \vec{x_i}; w)$$

→ Apply log to convert multiplication to Summation for easy Calculation
→ Max location doesn't change and will be Same for original and log of.

$$w = \underset{w}{\operatorname{argmax}} \sum_{i=1}^{n} \log P(y_i / \vec{x_i}, w)$$

$$= \underset{w}{\operatorname{argmax}} \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{1}{2\sigma^2}(w^T x_i - y_i)^2$$

$$= \underset{w}{\operatorname{argmax}} \sum_{i=1}^{n} -(w^T x_i - y_i)^2 = \underset{w}{\operatorname{argmin}} \sum_{i=1}^{n} (w^T x_i - y_i)^2$$

→ To make it interpretable ~~and~~ divide by n, so that it is loss per sample.

MAP:

$$p(w \mid y_1x_1\, y_2x_2 \cdots y_nx_n) = \frac{p(y_1x_1 \cdots y_nx_n \mid w)\, p(w)}{p(y_1x_1 \cdots y_nx_n)}$$ ← Prior

$p(w) \sim N(0, \tau^2 I)$

$\vec{w}_0$ in the case of transfer learning

$$w = \underset{w}{\arg\max}\; p(w \mid y_1x_1 \cdots y_nx_n)$$

$$= \underset{w}{\arg\max}\; p(y_1x_1 \cdots y_nx_n \mid w)\, p(w)$$

$$= \underset{w}{\arg\max} \left[\prod_{i=1}^{n} p(y_i x_i \mid w)\right] p(w)$$

$$= \underset{w}{\arg\max} \left[\prod_{i=1}^{n} p(y_i \mid x_i, w)\, \underbrace{p(x_i \mid w)}_{\text{constant}}\right] p(w)$$

$$= \underset{w}{\arg\max} \left[\prod_{i=1}^{n} p(y_i \mid x_i, w)\right] p(w)$$

Apply log

$$= \underset{w}{\arg\max} \sum_{i=1}^{n} \log p(y_i \mid x_i, w) + \log p(w)$$

$$= \underset{w}{\arg\min}\; \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i^T w - y_i)^2 + \frac{1}{2\tau^2} w^T w$$

$$= \underset{w}{\arg\min}\; \frac{1}{n} \sum_{i=1}^{n} (x_i^T w - y_i)^2 + \lambda \|w\|_2^2$$

$$\lambda = \frac{\sigma^2}{n\tau^2}$$

$$\frac{\partial L}{\partial w} = \cancel{2}\, XX^T w - XY^T + \lambda w$$

$$= XX^T w - XY^T + \lambda w$$

$$0 = XX^T w - XY^T + \lambda w$$

$d\times n \quad n\times d$

$d\times d \cdot d\times 1$

$d\times n \quad n\times 1$

$$(XX^T - \lambda I)\, w = XY^T$$

$$\boxed{w = (XX^T - \lambda I)^{-1} XY^T}$$ → closed form solution for Ridge regression / L_2 normalisation