MACHINE LEARNING

Homework Week 5

October 6, 2021

Nguyen Son Tung

---

1. Posterior

<u>Answer</u>

Bayes theorem

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}$$

$$\Leftrightarrow posterior = \frac{likelihood \times prior}{evidence}$$

$$\Rightarrow p(\mathbf{w}|\mathbf{x}, \mathbf{t}, \alpha, \beta) = \frac{p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta)p(\mathbf{w}|\alpha)}{p(\mathbf{x}, \mathbf{t}, \alpha, \beta)}$$

$p(\mathbf{w}|\mathbf{x}, \mathbf{t}, \alpha, \beta)$ is a posterior. While likelihood is given the parameter how the parameter fit the data, posterior is given the data, what is the probability of parameter. In the posterior, we also include our belief.

We expect to maximize the posterior to find $\mathbf{w}$.

$$p(\mathbf{w}|\mathbf{x}, \mathbf{t}, \alpha, \beta) \propto p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta)p(\mathbf{w}|\alpha)$$

Because $p(\mathbf{x}, \mathbf{t}, \alpha, \beta)$ is dependent of $\mathbf{w}$
Suppose $p(\mathbf{w}|\alpha)$ is a normal distribution. We have

$$p(\mathbf{w}|a) = \mathcal{N}(\mathbf{w}|0, \alpha^{-1}I)$$

$$p(\mathbf{t}|\mathbf{x}, \mathbf{w}) = \mathcal{N}(\mathbf{t}_n|y(\mathbf{x}_n, \mathbf{w}), \beta^{-1})$$

So

$$p(\mathbf{w}|\mathbf{x}, \mathbf{t}, \alpha, \beta)$$

$$\propto p(\mathbf{t}|\mathbf{x}, \mathbf{w}, \beta)p(\mathbf{w}|\alpha)$$

$$= \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\beta^{-1}}} exp\left\{ - \frac{\mathbf{t}_n - y(\mathbf{x}_n, \mathbf{w})^2}{2\beta^{-1}} \right\}$$

$$\times \frac{1}{(2\pi)^D |\alpha^{-1} I|} exp\left\{ - \frac{1}{2}\mathbf{w}^T (\alpha^{-1} I)^{-1} \mathbf{w} \right\}$$

$$\Rightarrow \log p(\mathbf{w}|\mathbf{x}, \mathbf{t}, \alpha, \beta)$$

$$\propto \ -\frac{\beta}{2} \sum_{n=1}^{n} \{y(\mathbf{x}_n, \mathbf{w})^2 - \mathbf{t}_n\}^2 - \frac{\alpha}{2}\mathbf{w}^T\mathbf{w}$$

we find that the maximum of the posterior is given by the minimum of

$$\frac{\beta}{2} \sum_{n=1}^{n} \{y(\mathbf{x}_n, \mathbf{w})^2 + \mathbf{t}_n\}^2 - \frac{\alpha}{2}\mathbf{w}^T\mathbf{w}$$

or we minimize

$$Q = ||X\mathbf{w} - \mathbf{t}||_2^2 + \lambda \mathbf{w}^T\mathbf{w} \qquad with \ \lambda = \frac{\alpha}{\beta}$$

$$\nabla Q_{\mathbf{w}} = X^T(X\mathbf{w} - t) + 2\lambda\mathbf{w}$$

$$\Rightarrow \mathbf{w} = (X^T X + \lambda I)^{-1} X^T t$$