DSEB K61, MFE, NEU

Machine Learning Assignment #5

Tô Đức Anh | 11196328

October 2021

1 Problems

1. Transform posterior to LaTeX, from $p(w|D)->w=(X^TX+\alpha*I)^{-1}X^Tt$

2 Answers

2.1 Question 1

We have the formula for Bayes theorem:

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

With posterior, base on Bayes theorem we have:

$$posterior=\frac{likelihood*prior}{evidence}$$

with A as w, and B as Data, we have

$$p(D|w)=\frac{p(w|D)p(w)}{p(D)}$$

but this is non sense, since we cant calculate(?) the probability of D given w, since we don't know w yet, so we transform it, we have that:

$$p(D|w)=p(w|D)$$

So that, we have:

$$p(w|D)=\frac{p(D|w)p(w)}{p(D)}$$

with
$p(w|D)$ as Posterior,
$p(D|w)p(w)$ as Prior,
$p(D)$ as Evidence

$$p(w|x,t)=\frac{p(t|x,w)p(w|\alpha)}{p(D)}$$

Assume that the prior $p(D|w)p(w)$ follow the Normal distribution, so that we have:

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

And

$$p(t|x,w)=\prod_{i=1}^{n}p(t_i|x_i,w)=\prod_{i=1}^{n}\mathcal{N}(t_i|y(x_i,w),\beta^{-1})$$

So, to maximize the posterior we have to maximize

$$p(w,\alpha)*p(t_i|x_i,w)$$

or maximize:

$$\mathcal{N}(w|0, a^{-1}I) * \prod_{i=1}^{n} \mathcal{N}(t_i|y(x_i, w), \beta^{-1})$$

take the logarithm, which means we will maximize

$$\log_e(\mathcal{N}(w|0, a^{-1}I) * \prod_{i=1}^{n} \mathcal{N}(t_i|y(x_i, w), \beta^{-1}))$$

$$\log_e p(t|x, w) + \log_e(p(w|\alpha))$$

$$\sum_{i=1}^{n} \log_e(\mathcal{N}(y(x_i, w)), \beta^{-1}) + \log_e(p(w|\alpha))$$

$$\sum_{i=1}^{n} \log_e(\mathcal{N}(y(x_i, w)), \beta^{-1}) + \log_e(p(w|\alpha))$$

$$\sum_{i=1}^{n} (\frac{1}{\beta^{-1}\sqrt{2\pi}}) e^{\frac{-(t_i - y(x_i, w))^2 \beta}{2}} \log(\frac{1}{\sqrt{2\pi^D} |\alpha^{-1} I|} e^{\frac{-1}{2} w^T (\alpha^{-1} I)^{-1} w}$$

note: more explanation on why $|\alpha|^{-1} I_D$ disappear can be watched in the lecture video.

$$L = -\frac{\beta}{2} \sum_{i=1}^{n} (t_i - y(x_i, w))^2 + -\frac{1}{2} \alpha w^T w$$

$$-\frac{\beta}{2} \sum_{i=1}^{n} (t_i - y(x_i, w))^2 + \frac{\alpha}{\beta} w^T w$$

Since we need to maximize the Posterior, so we need to minimize L.
We take the derivative.
Let $\frac{\alpha}{\beta} = \lambda$

$$-\frac{\beta}{2} \sum_{i=1}^{n} (t_i - y(x_i, w))^2 + \lambda w^T w$$

$$L = ||Xw - t||_2^2 + \lambda ||w||_2^2$$

$$\frac{dL}{dw} = 2X^T(Xw - t) + 2\lambda w = 0$$

$$w(X^T X + \lambda I_n) = X^T t$$

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