Introduction to Statistical Machine Learning Homework 4

Yota Toyama

November 30, 2016

1.

$$\mathbf{w}^* = \underset{\mathbf{w}}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} (y_i - \mathbf{w} \cdot \phi(x_i))^2 \right\}$$
$$\text{s.t. } \sum_{j=1}^{d} w_j^2 \leq \tau$$

Using Lagrange's multiplier method,

$$\min_{\lambda} \max_{\mathbf{w}} - \sum_{i=1}^{N} (y_i - \mathbf{w} \cdot \phi(x_i))^2 + \lambda(\tau - \sum_{j=1}^{d} w_j^2)$$
$$\max_{\lambda} \min_{\mathbf{w}} \sum_{i=1}^{N} (y_i - \mathbf{w} \cdot \phi(x_i))^2 - \lambda(\tau - \sum_{j=1}^{d} w_j^2)$$
$$\frac{\partial}{\partial \mathbf{w}} \left\{ - \sum_{i=1}^{N} (y_i - \mathbf{w} \cdot \phi(x_i))^2 + \lambda(\tau - \sum_{j=1}^{d} w_j^2) \right\} = 0$$
$$2 \sum_{i=1}^{N} \phi(x_i)(y_i - \mathbf{w} \cdot \phi(x_i)) - 2\lambda \mathbf{w} = 0$$
$$\sum_{i=1}^{N} \phi(x_i)(y_i - \mathbf{w} \cdot \phi(x_i)) - \lambda \mathbf{w} = 0$$

The equasion above is the same as the derivative of L2 regularized squared loss in regression.
$\therefore$ If a proper τ which makes λs of both problems same is found, the w^*s of both problems are same.

2.

$$
\begin{aligned}
\hat{y} &= \underset{c}{\operatorname{argmax}} \{\log p(\mathbf{x}, y = c)\} \\
&= \underset{c}{\operatorname{argmax}} \{\log p(\mathbf{x}|y = c) + \log p(y = c)\} \\
&= \underset{c}{\operatorname{argmax}} \{\log p(\mathbf{x}|y = c)\}
\end{aligned}
$$

$$P(\mathbf{x}|y = c; \theta) = \prod_{j=1}^{d} \theta_j^{x_j} (1 - \theta_j)^{1 - x_j}$$

$$\log P(\mathbf{x}|y = c; \theta) = \sum_{j=1}^{d} x_j \log \theta_j + (1 - x_j) \log(1 - \theta_j)$$

$$\frac{\partial \log P(X|\mathbf{y} = c; \theta)}{\partial \theta} = 0$$

$$\frac{x_j}{\theta_j} - \frac{1 - x_j}{1 - \theta_j} = 0$$

$$\theta_j = \frac{1}{N} \sum_{i=1}^{N} x_j$$

References

[1] Christopher M. Bishop, Pattern Recognition and Machine Learning

[2] Discussion with Tomoki Tsujimura and Bowen Shi