Introduction to Statistical Machine Learning
Homework 4

Yota Toyama

November 30, 2016

1.

$$\mathbf{w}^* = \operatorname*{argmin}_{\mathbf{w}} \left\{ \sum_{i=1}^{N} (y_i - \mathbf{w} \cdot \phi(x_i))^2 \right\}$$

$$\text{s.t.} \ \sum_{j=1}^{d} w_j^2 \leq \tau$$

Using Lagrange's multiplier method,

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

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

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

$$2\phi(x_i)(y_i - \mathbf{w} \cdot \phi(x_i)) - 2\lambda\mathbf{w} = 0 \phi(x_i)(y_i - \mathbf{w} \cdot \phi(x_i)) - \lambda\mathbf{w} = 0$$

2.

$$\hat{y} = \operatorname*{argmax}_{c} \left\{ \log p(\mathbf{x}, y = c) \right\}$$

$$= \operatorname*{argmax}_{c} \left\{ \log p(\mathbf{x}|y = c) + \log p(y = c) \right\}$$

$$= \operatorname*{argmax}_{c} \left\{ \log p(\mathbf{x}|y = c) \right\}$$

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

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

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

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

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

References

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

[2] Discussion with Tomoki Tsujimura and Bowen Shi