AMIT YADAV

Assignment #2

CS-E4830 Kernel Methods in ML

① We know that

$$k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$$

Now,

$$k_c(x_i, x_j) = \langle \phi_c(x_i), \phi_c(x_j) \rangle$$

$$= \langle (\phi(x_i) - \overline{\phi(x)}), (\phi(x_j) - \overline{\phi(x)})$$

where $\overline{\phi(x)} = \frac{1}{N} \sum_{p=1}^{N} \phi(x_p)$

$$= \langle \phi(x_i), \phi(x_j) \rangle - \langle \phi(x_j), \overline{\phi(x)} \rangle - \langle \phi(x_i), \overline{\phi(x)} \rangle$$
$$+ \langle \overline{\phi(x)}, \overline{\phi(x)} \rangle$$

$$= k(x_i, x_j) - \langle \phi(x_j), \frac{1}{N} \sum_{p=1}^{N} \phi(x_p) \rangle$$
$$- \langle \phi(x_j), \frac{1}{N} \sum_{p=1}^{N} \phi(x_p) + \frac{1}{N^2} \langle \sum_{p=1}^{N} \phi(x_p), \sum_{p=1}^{N} \phi(x_p) \rangle$$

$$= k(x_i, x_j) - \frac{1}{N} \sum_{p=1}^{N} \langle \phi(x_i), \phi(x_p)) - \frac{1}{N} \sum_{p=1}^{N} \langle \phi(x_j, \phi(x_p) \rangle$$
$$+ \frac{1}{N^2} \sum_{p=1}^{N} \sum_{q=1}^{N} \langle \phi(x_p), \phi(x_q) \rangle$$

$$= K(x_i, x_j) - \frac{1}{N}\sum_{p=1}^{N} K(x_i, x_p) - \frac{1}{N}\sum_{p=1}^{N} K(x_j, x_p)$$

$$+ \frac{1}{N^2}\sum_{p=1}^{N}\sum_{q=1}^{N} K(x_p, x_q)$$

$$= RHS$$

(2)

(1) $P(y=c_1 \mid x=\hat{x}) = \dfrac{P(c_1) \cdot P(x=\hat{x} \mid c_1)}{P(x=\hat{x})}$

$= \dfrac{P(x=\hat{x} \wedge y=c_1)}{P(x=\hat{x})}$

$= \dfrac{P(\hat{x}, c_1)}{P(\hat{x}, c_1) + P(\hat{x}, c_2)}$

(2) As we can see from the plot that coloured regions give us the total loss, if $\hat{x}$ is the decision boundary. The loss will be minimized if $x = x_0$ is the decision boundary.

That is, for any x, we assign the class for which joint-probability $P(x, c)$ is maximum.

$\therefore P(\text{Misclassification error}) = \min(P(x,c_1), P(x,c_2))$

Integrating over all x, we get

$$P(m.m.E) = \int_{x \in X} \min(P(x,c_1), P(x,c_2))\, dx$$

$$\leq \int_{x \in X} (P(x,c_1), P(x,c_2))^{1/2}\, dx$$

3

$$E[e^{\lambda\varepsilon}] = E\left[1 + \lambda\varepsilon + \frac{(\lambda\varepsilon)^2}{2!} + \frac{(\lambda\varepsilon)^3}{3!} + \cdots + \frac{(\lambda\varepsilon)^n}{n!} + \cdots\right]$$

$$= E[1] + E[\lambda\varepsilon] + \cdots + E\left[\frac{(\lambda\varepsilon)^n}{n!}\right] + \cdots$$

$$= 1 + \lambda E[\varepsilon] + \frac{\lambda^2}{2!}E[\varepsilon^2] + \cdots + \frac{\lambda^n}{n!}E[\varepsilon^n] + \cdots$$

$$= 1 + \frac{\lambda^2}{2!} + \frac{\lambda^4}{4!} + \cdots + \frac{\lambda^{2n}}{(2n)!} + \cdots$$

$$\left(\because \begin{array}{ll} E[\varepsilon^n] = 0 & \text{if } n \text{ is odd} \\ = 1 & \text{if } n \text{ is even} \end{array}\right)$$

Now, RHS $= e^{\lambda^2/2}$

$$= 1 + \left(\frac{\lambda^2}{2}\right) + \left(\frac{\lambda^2}{2}\right)^2 \times \frac{1}{2!} + \cdots + \left(\frac{\lambda^2}{2}\right)^n \times \frac{1}{n!} + \cdots$$

Comparing n-th term of LHS and RHS

$$\frac{\lambda^{2n}}{(2n)!} \leq \frac{\lambda^{2n}}{2^n n!} \quad \text{for all } n \geq 1$$

$\therefore$ LHS $\leq$ RHS

$$\Rightarrow \boxed{E[e^{\lambda\varepsilon}] \leq e^{\lambda^2/2}}$$

Scanned by CamScanner