CSC411H1 FALL 2018

HW 7

LINO LASTELLA 1001237654 LASTELL1

②

a) Let $k_1(x,x') = \psi_1(x)^T \psi_1(x')$, $k_2(x,x') = \psi_2(x)^T \psi_2(x')$

Then $k_s(x,x') = k_1(x,x') + k_2(x,x')$

$$= \psi_1(x)^T \psi_1(x') + \psi_2(x)^T \psi_2(x')$$

$$= \begin{pmatrix} \psi_1(x) \\ \psi_2(x) \end{pmatrix}^T \begin{pmatrix} \psi_1(x') \\ \psi_2(x') \end{pmatrix}$$

$$= \psi_s(x)^T \psi_s(x) \quad \left[\text{Substitution } \psi_s(x) = \begin{pmatrix} \psi_1(x) \\ \psi_2(x) \end{pmatrix} \right]$$

Feature map is therefore $\psi_s(x) = \begin{pmatrix} \psi_1(x) \\ \psi_2(x) \end{pmatrix}$. $\square$

b) Let $k_1(x,x') = \psi_1(x)^T \psi_1(x')$, $k_2(x,x') = \psi_2(x)^T \psi_2(x')$

Then, $k_p(x,x') = k_1(x,x') k_2(x,x')$

$$= \psi_1(x)^T \psi_1(x') \ \psi_2(x)^T \psi_2(x')$$

Now assume $\psi_1(x) = \begin{pmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_d(x) \end{pmatrix}$, $\psi_2(x) = \begin{pmatrix} g_1(x) \\ g_2(x) \\ \vdots \\ g_{d'}(x) \end{pmatrix}$

then,

$$k_p(x, x') = \sum_{i=1}^{d} f_i(x) f_i(x') \sum_{j=1}^{d'} g_j(x) g_j(x')$$

$$= \sum_{i=1}^{d} \sum_{j=1}^{d'} f_i(x) g_j(x) f_i(x') g_j(x')$$

$$= \sum_{i=1}^{d} \sum_{j=1}^{d'} h_{ij}(x) h_{ij}(x')$$

$$= \psi_p(x)^T \psi_p(x'), \quad \text{where}$$

$$\psi_p(x) = \begin{pmatrix} h_{1\cdot 1}(x) \\ h_{2\cdot 1}(x) \\ \vdots \\ h_{d \cdot d'}(x) \end{pmatrix} = \begin{pmatrix} f_1(x) g_1(x) \\ f_2(x) g_1(x) \\ \vdots \\ f_d(x) g_{d'}(x) \end{pmatrix}$$

$\square$

①

2) $J(w) = \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(y^{(i)}, t^{(i)}) + \frac{\lambda}{2} \|w\|^2$

Let S = row space of Ψ = span $\{ \psi(x^{(i)})^T, i = 1, 2, \cdots, N \}$.

Then, $w = w_S + w_\perp$, where w_S is the projection of w onto S and $w_\perp$ is orthogonal to S (Hint).

Let w^* be a minimizer for $J(w)$.

Then,

$$J(w^*) = \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(g(w^{*T} \psi(x)), t^{(i)}) + \frac{\lambda}{2} \|w^*\|^2$$

$$= \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(g(w_S^{*T} \psi(x)), t^{(i)}) + \frac{\lambda}{2} \|w_S^*\|^2 + \frac{\lambda}{2} \|w_\perp^*\|^2$$

$$\geq \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(g(w_S^{*T} \psi(x)), t^{(i)}) + \frac{\lambda}{2} \|w_S^*\|^2$$

$$= J(w_S^*).$$

This means that for every minimizer w^*, the projection of w^* onto S is actually the real global minimum.

In other words, the optimal weights lie in span $\{ \psi(x^{(i)})^T \}$, $i = 0, 1, \cdots, N$

□

b)

$$J(w) = \frac{1}{2N} \| t - \Psi w \|^2 + \frac{\lambda}{2} \| w \|^2$$

$$J(\alpha) = \frac{1}{2N} \| t - \Psi \Psi^T \alpha \|^2 + \frac{\lambda}{2} \| \Psi^T \alpha \|^2$$ [Substitution $w = \Psi^T \alpha$]

$$= \frac{1}{2N} \| t - K\alpha \|^2 + \frac{\lambda}{2} \| \Psi \cdot \alpha \|^2$$ dot-product

In order to find a candidate for minimizing this function, we can set its gradient equal to zero:

$$\nabla J(\alpha) = \frac{1}{2N} 2 (-K) (t - K\alpha) + \frac{\lambda}{2} 2 \Psi \ (\Psi \cdot \alpha)$$

$$= -\frac{1}{N} K (t - K\alpha) + \lambda \underbrace{\Psi \Psi^T}_{K} \alpha$$

$$= \frac{1}{N} K^2 \alpha - \frac{1}{N} K t + \lambda K \alpha = 0$$

$$\Rightarrow \left(\frac{1}{N} K - \lambda I\right) \alpha = \frac{1}{N} t$$ [First multiply both sides by K^{-1} then move $\frac{1}{N} t$ to RHS]

[Note: I = identity matrix of the same size as K].

$$\Rightarrow \alpha = \frac{1}{N} \left(\frac{1}{N} K - \lambda I\right)^{-1} t$$

The only candidate is therefore $\alpha_0 = \frac{1}{N}\left(\frac{1}{N}K - \lambda I\right)^{-1} t$.

To check that this is a real (global) minimum, we need to verify that $HJ(\alpha_0) > 0$ [the Hessian of J evaluated at α_0 is positive-definite]

$$HJ(\alpha_0) = HJ(\alpha)\Big|_{\alpha=\alpha_0} = \frac{1}{N}K^2 + \lambda K\Big|_{\alpha=\alpha_0} = \frac{1}{N}K^2 + \lambda K.$$

Since K is positive-definite by assumption and N, λ are both greater than zero, we can conclude that $HJ(\alpha_0) > 0$.

$$\therefore \quad \alpha_0 = \frac{1}{N}\left(\frac{1}{N}K - \lambda I\right)^{-1} t \quad \text{is the only global minimum}$$

$\square$