CSC411 A7

Zhihong Wang 1002095207

2018-12

Q1:

(a).

Let $S = span_i\{\psi(x^{(i)})\}$.

The optimal weights vector w^* can be decomposed as $w^* = w^*_S + w^*_\perp$, where w^*_S is the projection of w^* onto S, and $w^*_\perp$ is orthogonal to S.

Then, we know w^* lies in the row space of Ψ if and only if $w^*_\perp$ is 0.

Prove by contradiction:

Assume $w^*_\perp$ is not 0.

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

For $\jmath(w^*)$, we can get:

$$
\begin{aligned}
L(y^{(i)}, t^{(i)}) &= L(g(w^{*\top}\psi(x^{(i)})), t^{(i)}) \\
&= L(g((w^*_S + w^*_\perp)^\top\psi(x^{(i)})), t^{(i)}) \\
&= L(g(w^{*\top}_S\psi(x^{(i)}) + w^{*\top}_\perp\psi(x^{(i)})), t^{(i)}) \\
&= L(g(w^{*\top}_S\psi(x^{(i)})), t^{(i)})
\end{aligned}
$$

(Since $w^*_\perp$ is orthogonal to vectors in S, their dot product is 0)

For $\jmath(w^*_S)$, we can get:

$$L(y^{(i)}, t^{(i)}) = L(g(w^{*\top}_S\psi(x^{(i)})), t^{(i)})$$

So we know the $L(y^{(i)}, t^{(i)})$ for both $\jmath(w^*)$ and $\jmath(w^*_S)$ are the same.

Also, $\|w^*\|^2 = w^{*\top} w^*$

$$= (w_S^* + w_\perp^*)^\top (w_S^* + w_\perp^*)$$

$$= w_S^{*\top} w_S^* + w_S^{*\top} w_\perp^* + w_S^* w_\perp^{*\top} + w_\perp^{*\top} w_\perp^*$$

$\|w_S^*\|^2 = w_S^{*\top} w_S^*$

So we know $\|w^*\|^2 > \|w_S^*\|^2$

Then $\jmath(w^*) > \jmath(w_S^*)$, which is contradicted with w^* is the optimal weights.

Therefore, w^* must lie in the row space of Ψ.

(b).

$w = \Psi^\top \alpha, K = \Psi\Psi^\top$

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

$$= \frac{1}{2N}\|t - \Psi\Psi^\top\alpha\|^2 + \frac{\lambda}{2}\|\Psi^\top\alpha\|^2$$

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

$$= \frac{1}{2N}(t - K\alpha)^\top(t - K\alpha) + \frac{\lambda}{2}(\Psi^\top\alpha)^\top(\Psi^\top\alpha)$$

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

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

Let $A = \frac{K^\top K}{N} + \lambda K, b^\top = -\frac{t^\top K}{N}, b = -\frac{K^\top t}{N}, c = \frac{t^\top t}{2N}$

then $\jmath(w) = \frac{1}{2}\alpha^\top A\alpha + b^\top\alpha + c$

So $\alpha = -A^{-1}b = (\frac{K^\top K}{N} + \lambda K)^{-1}\frac{K^\top t}{N}$

Q2:

$$k_1(x,x^{'}) = \psi_1(x)^\top \psi_1(x^{'}), k_2(x,x^{'}) = \psi_2(x)^\top \psi_2(x^{'})$$

(a).

$$\begin{aligned} k_S(x,x^{'}) &= k_1(x,x^{'}) + k_2(x,x^{'}) \\ &= \psi_1(x)^\top \psi_1(x^{'}) + \psi_2(x)^\top \psi_2(x^{'}) \\ &= (\psi_1(x)\psi_2(x)) \begin{bmatrix} \psi_1(x^{'}) \\ \psi_2(x^{'}) \end{bmatrix} \\ &= \psi_S(x)^\top \psi_S(x^{'}) \end{aligned}$$

Therefore, the feature map $\psi_S = \begin{bmatrix} \psi_1(x) \\ \psi_2(x) \end{bmatrix}$

(b).

$$\begin{aligned} k_P(x,x^{'}) &= k_1(x,x^{'})k_2(x,x^{'}) \\ &= \textstyle\sum_{i=1}^{n} \psi_1^{(i)}(x)\psi_1^{(i)}(x^{'}) \sum_{j=1}^{m} \psi_2^{(j)}(x)\psi_2^{(j)}(x^{'}) \\ &= \textstyle\sum_{i=1}^{n} \sum_{j=1}^{m} (\psi_1^{(i)}(x)\psi_2^{(j)}(x))(\psi_1^{(i)}(x^{'})\psi_2^{(j)}(x^{'})) \\ &= \textstyle\sum_{k=1}^{nm} \psi_P^{(k)}(x)\psi_P^{(k)}(x^{'}) \\ &= \psi_P(x)^\top \psi_P(x^{'}) \end{aligned}$$

Therefore, the feature map $\psi_P = \begin{bmatrix} \psi_P^{(1)}(x) \\ \psi_P^{(2)}(x) \\ ... \\ \psi_P^{(nm)}(x) \end{bmatrix}$

$\psi_P^{(1)}(x) = \psi_1^{(1)}(x)\psi_2^{(1)}(x), \psi_P^{(2)}(x) = \psi_1^{(1)}(x)\psi_2^{(2)}(x), ...,$

$\psi_P^{(m)}(x) = \psi_1^{(1)}(x)\psi_2^{(m)}(x), ..., \psi_P^{(nm)}(x) = \psi_1^{(n)}(x)\psi_2^{(m)}(x)$