Vele Tosevski

1002174657

CSC 411

April 3, 2019

HW7 Write-Up

1. a). We have the following formula for minimization:

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

where $\|w\|^2 = w^T w$ and L is a loss function with

$$y^{(i)} = g(z^{(i)}) = g(w^T \psi(x^{(i)})) \quad (2)$$

Plugging (2) into (1),

$$\mathcal{J}(w) = \frac{1}{N}\sum_{i=1}^{N} L(g(w^T\psi(x^{(i)})), t^{(i)}) + \frac{\lambda}{2} w^T w \quad (3)$$

We can decompose w according to the hint provided.

$W = w_\psi + w_\perp$ (4) where w_ψ is a projection of w onto the ψ subspace and $w_\perp$ is perpendicular to the ψ subspace.

$\therefore$ (4) $\rightarrow w^T w$

$$\begin{aligned}
w^T w &= (w_\psi + w_\perp)^T (w_\psi + w_\perp) \\
&= (w_\psi^T + w_\perp^T)(w_\psi + w_\perp) \\
&= w_\psi^T w_\psi + \cancel{w_\psi^T w_\perp}^{0} + \cancel{w_\perp^T w_\psi}^{0} + w_\perp^T w_\perp \\
&= w_\psi^T w_\psi + w_\perp^T w_\perp \quad (5)
\end{aligned}$$

since w_ψ is ortho. to $w_\perp$

$\therefore$ (4) $\rightarrow L(g(w^T\psi(x^{(i)})), t^{(i)})$

$$L(\dots) = L\left[g((w_\psi + w_\perp)^T \psi(x^{(i)})), t^{(i)}\right]$$

$$= L\left[g(w_\psi^T\psi(x^{(i)}) + \underbrace{w_\perp^T\psi(x^{(i)})}_{0}), t^{(i)}\right]$$

$$= L\left[g(w_\psi^T\psi(x^{(i)})), t^{(i)}\right]$$ since $w_\perp$ is ortho. to every row in $\psi(x^{(i)})$

(6)

$\therefore$ (5), (6) $\rightarrow$ (3)

$$J(w) = \frac{1}{N}\sum_{i=1}^{N} L\left(g(w_\psi^T\psi(x^{(i)})), t^{(i)}\right) + \frac{\lambda}{2}\left(w_\psi^T w_\psi + w_\perp^T w_\perp\right)$$

$$\geq \frac{1}{N}\sum_{i=1}^{N} L\left(g(w_\psi^T\psi(x^{(i)})), t^{(i)}\right) + \frac{\lambda}{2}\left(w_\psi^T w_\psi\right)$$

$$= J(w_\psi)$$

Since $J(w) \geq J(w_\psi)$, we know w_ψ is a more optimal result. w_ψ has a linear combination of $\psi(x^{(1)}) \dots \psi(x^{(N)})$ meaning it is in the row space of Ψ.
So the optimal weights must lie in row space of Ψ. $\blacksquare$

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

Plugging in $w = \Psi^T\alpha$ for some α, ...

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

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

$$= \frac{1}{2N}\left(t^T t - t^T\Psi\Psi^T\alpha - \Psi^T\Psi\alpha^T t + \Psi^T\Psi\alpha^T\Psi\Psi^T\alpha\right) + \frac{\lambda}{2} w^T w$$

$$= \frac{1}{2N}\left(t^T t - t^T\Psi\Psi^T\alpha - \underline{\Psi^T\Psi\alpha^T t} + \Psi^T\Psi\alpha^T\Psi\Psi^T\alpha\right) + \frac{\lambda}{2}\Psi\alpha^T\Psi^T\alpha$$

$$2N \cdot J(\alpha) = t^T t - t^T K\alpha - K\alpha^T t + K\alpha^T K\alpha + N\lambda \cdot \alpha^T K\alpha$$

$$= t^T t - 2t^T K\alpha + \alpha^T KK\alpha + N\lambda \cdot \alpha^T K\alpha$$

$$= \alpha^T KK\alpha + N\lambda \cdot \alpha^T K\alpha - 2t^T K\alpha + t^T t$$

$$= \alpha^T (KK\alpha + N\lambda K\alpha) - 2t^T K\alpha + t^T t$$

$$= \alpha^T (KK + N\lambda \cdot K)\alpha - 2t^T K\alpha + t^T t$$

$$N \cdot J(\alpha) = \frac{1}{2}\alpha^T \underbrace{(KK + N\lambda \cdot K)}_{=A}\alpha - \underbrace{(t^T K)}_{=-b^T}\alpha + \underbrace{\frac{1}{2}t^T t}_{=c}$$

$\therefore$ Using the assumed hint from handout that min of

$J(w) \Rightarrow \alpha = -A^{-1}b$,

$$\underset{\alpha}{\arg\min}\left[\frac{1}{2}\alpha^T A\alpha + b^T\alpha + c\right] = -A^{-1}b$$

$$= -(KK + N\lambda \cdot K)^{-1}(-t^T K)^T$$

$$\alpha = \boxed{(KK + N\lambda K)^{-1}(K^T t)}$$

2.a). We have,

$$k_1(x,x') = \psi_1(x)^T \psi_1(x') \quad (1)$$

$$k_2(x,x') = \psi_2(x)^T \psi_2(x') \quad (2)$$

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

$$= \psi_1(x)^T \psi_1(x') + \psi_2(x)^T \psi_2(x') \quad (3)$$

We need, $k_s(x,x') = \psi_s(x)^T \psi_s(x')$ (4) for some ψ_s.

Now, (3) $\rightarrow$ (4)

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

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

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

$$= \left(\psi_s(x) \right)^T \left(\psi_s(x') \right)$$

$$\therefore \boxed{\psi_s(x) = \begin{pmatrix} \psi_1(x) \\ \psi_2(x) \end{pmatrix}}$$

b). We have,

$$k_1(x,x') = \psi_1(x)^T \psi_1(x') \quad (1)$$

$$k_2(x,x') = \psi_2(x)^T \psi_2(x') \quad (2)$$

$$k_p(x,x') = k_1(x,x')\, k_2(x,x') = \psi_1(x)^T \psi_1(x')\, \psi_2(x)^T \psi_2(x') \quad (3)$$

We need,

$$k_p(x,x') = \psi_p(x)^T \psi_p(x') \quad (4) \text{ for some } \psi_p$$

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

$$= \left[\sum_{i=1}^{A} \left(f_i(x)\, f_i(x')\right)\right]\left[\sum_{j=1}^{B} \left(g_j(x)\, g_j(x')\right)\right]$$

$$= \sum_{i=1}^{A} \sum_{j=1}^{B} \left(f_i(x)\, g_j(x)\right)\left(f_i(x')\, g_j(x')\right)$$

$$= \sum_{i=1}^{A} \sum_{j=1}^{B} \left(h_{ij}(x)\, h_{ij}(x')\right)$$

$$= \psi_p(x)^T \psi_p(x')$$

$\therefore$

$$\boxed{\psi_p(x) = \begin{pmatrix} h_{11}(x) \\ h_{21}(x) \\ h_{31}(x) \\ \vdots \\ h_{AB}(x) \end{pmatrix} = \begin{pmatrix} f_1(x)\, g_1(x) \\ f_2(x)\, g_1(x) \\ f_3(x)\, g_1(x) \\ \vdots \\ f_A(x)\, g_B(x) \end{pmatrix}}$$