CSC 411 Fall 2018

Machine Learning and Data Mining

Homework 7

Family name: Bhatia

Given name: Pooja

Q1 Representer Theorem

Solution part (a)

$$Z = \omega^T \varphi(x) \quad — ②$$

$$y = g(z) \quad — ③$$

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

$$\Psi = \begin{pmatrix} \psi(x^{(1)})^T \\ \vdots \\ \psi(x^{(N)})^T \end{pmatrix} \quad \text{(feature matrix)}$$

To minimize the loss plus regularization project ω on a subspace

span $\{ \psi(x^{(i)}) : 1 \leq i \leq N \}$

ω_ψ is the component along the subspace

$\omega_\perp$ is the component orthogonal / perpendicular to subspace.

$$\omega = \omega_\psi + \omega_\perp \quad \text{(Decomposition of } \omega)$$

Now,
The regularizer term in equation 1

$$||w||^2 = ||w_\psi||^2 + ||w_\perp||^2 \geq ||w_\psi||^2$$ (from Pythagoras)

then,

$$\frac{\lambda}{2}\left(||w||^2\right) \geq \frac{\lambda}{2}\left(||w_\psi||^2\right)$$

hence This term is minimized for $\boxed{W = W_\psi}$ ∴ Regularizer term is minimized ~~for~~ when

$$\boxed{w = w_\psi}$$

The objective is to minimize loss and regulizer in equation 1 basically

$$w^* \in \arg\min\left(\frac{1}{N}\sum_{i=1}^{N} h(y^{(i)}, t^{(i)}) + \frac{\lambda}{2}||w||^2\right)$$

The individual loss terms in this would be: from 1, 2, 3 equation and taking Feature matrix.

$$\langle w, \psi(x^{(i)})\rangle = \langle w_\psi, \psi(x^{(i)})\rangle + \langle w_\perp, \psi(x^{(i)})\rangle$$

$$\langle \omega, \psi(x^{(i)}) \rangle = \langle \omega_\psi, \psi(x^{(i)}) \rangle + \langle \omega_\perp, \psi(x^{(i)}) \rangle$$

$$= \langle \omega_\psi, \psi(x^{(i)}) \rangle$$

Since $\langle \omega_\perp, \psi(x^{(i)}) \rangle = 0$ for all $i = 1 \dots N$ as $\psi(x^{(i)})$ belongs to subspace, and $\omega_\perp$ is perpendicular to subspace.

$\therefore$ This implies that loss $L(\cdot\cdot)$ only depends on component of ω that lies of the subspace. Hence to minimize loss $\|\omega_\perp\|$ is taken as zero.

$\therefore$ The optimal weights lie in row space of ψ.

Solution b

$$w = \varphi^T \alpha \quad -①$$

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

Substituting w in cost fuction (① in ②)

$$J(\alpha) = \frac{1}{2N} \|t - \varphi \varphi^T \alpha\|^2 + \frac{\lambda}{2} \|\varphi^T \alpha\|^2 \quad -③$$

given,

GRAM MATRIX $K = \varphi \varphi^T$

substitute in equation ③

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

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

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

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

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

$$J(\alpha) = \frac{1}{2N}\left(\|t\|^2 - 2t^TK\alpha + K^TK\alpha^T\alpha\right) + \frac{\lambda}{2}\left(K\alpha^T\alpha\right)$$

$$= \frac{\|t\|^2}{2N} - \frac{2t^TK\alpha}{2N} + \frac{K^TK\alpha^T\alpha}{2N} + \frac{\lambda}{2}\left(K\alpha^T\alpha\right)$$

$$J(\alpha) = \frac{\alpha^T\alpha}{2}\left(\frac{K^TK}{N} + \lambda K\right) - \left(\frac{2t^TK}{2N}\right)\alpha + \frac{\|t\|^2}{2N} \quad \text{— (4)}$$

$J(\alpha)$ the cost function is a quadratic function of the from

$$\frac{1}{2}\alpha^TA\alpha + b^T\alpha + C \quad \text{— (5)}$$

$A \rightarrow$ positive definate matrix A

$b \rightarrow$ vector

$c \rightarrow$ constant

Comparing (4) and (5) and using from assignment the minimum of such an equation (quadratic function) is given by $\boxed{\alpha = -A^{-1}b}$

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

$$\boxed{\alpha = \left(\frac{K^TK}{N} + \lambda K\right)^{-1}\left(\frac{t^TK}{N}\right)}$$

Ans

Q2 Compositional Kernels

Solution Part (a)

$$K_1(x, x') = \psi_1(x)^T \psi_1(x') \quad \text{— (1)}$$

$$K_2(x, x') = \psi_2(x)^T \psi_2(x') \quad \text{— (2)}$$

$$K_s(x, x') = K_1(x, x') + K_2(x, x')$$

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

(from (1) and (2))

$$= (\psi_1(x), \psi_2(x)) \begin{pmatrix} \psi_1(x') \\ \psi_2(x') \end{pmatrix} \quad \text{— (3)}$$

$$K_s(x, x') = \psi_s(x)^T \psi_s(x') \quad \text{— (4)}$$

comparing (3) and (4)

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

$$\boxed{\psi_s(x) = \begin{pmatrix} \psi_1(x) \\ \psi_2(x) \end{pmatrix}} \quad \underline{\text{Ans}}$$

$\boxed{\text{Solution part (b)}}$

$$K_1(x, x') = \psi_1(x)^T \psi_1(x') \quad — ①$$

$$K_2(x, x') = \psi_2(x)^T \psi_2(x') \quad — ②$$

(from sub ① and ②)

$$K_p(x, x') = K_1(x, x')\, K_2(x, x')$$

$$= \sum_{i=1}^{n} \psi_{1i}(x)\, \psi_{1i}(x') \sum_{j=1}^{m} \psi_{2j}(x)\, \psi_{2j}(x')$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} (\psi_{1i}(x)\, \psi_{2j}(x))\, (\psi_{1i}(x')\, \psi_{2j}(x'))$$

$$= \sum_{k=1}^{nm} \psi_{12k}(x)\, \psi_{12k}(x') \quad — ③$$

$$K_p(x, x') = \psi_p(x)^T \psi_p(x') \quad — ④$$

ψ_p is a feature map:

$\psi_p(x) = \psi_1(x) \times \psi_2(x)$ (cartesian product)

(By comparing 3 and 4)

$$\boxed{\psi_p(x) = \psi_1(x) \times \psi_2(x)}$$

Ans.

In the above solution, I have taken $\psi_1(x)$ to be n dimensional vector and $\psi_2(x)$ to be m dimensional vector where $\psi_{1i}(x)$ is the ith feature value under feature map ψ_1 and $\psi_{2j}(x)$ is the j^{th} feature value under feature map ψ_2. We can also solve the same by taking limit/range to ∞ instead of n and m. The same is shown below,

$$K_p(x, x') = k_1(x, x')\, k_2(x, x')$$

$$= \left(\sum_{i=1}^{\infty} \psi_{1i}(x)\, \psi_{1i}(x')\right)\left(\sum_{j=1}^{\infty} \psi_{2j}(x)\, \psi_{2j}(x')\right)$$

$$= \sum_{i,j} \psi_{1i}(x)\psi_{1i}(x')\psi_{2j}(x)\psi_{2j}(x') \rightarrow 5$$

$$\boxed{\psi_p(x) = \psi_1(x) \times \psi_2(x)}$$

(from comparing 4 and 5)

$\rightarrow$ Ans