CSC411 Fall 2018 – Homework 7 Solution

1 Representer Theorem

Part a)

Let $\mathbf{w} = \mathbf{w}_\Psi + \mathbf{w}_\perp$, in which $\mathbf{w}_\Psi$ is the projection onto Ψ. Note that $z = \mathbf{w}_\Psi^T \Psi + \mathbf{w}_\perp^T \Psi = \mathbf{w}_\Psi^T \Psi$ due to orthogonality. Therefore, the minimizer of expected loss $\sum_{i=1}^N \mathcal{L}(y^{(i)}, t^{(i)})$ does not depend on $\mathbf{w}_\perp$. **(+1pt)**

On the other hand, $\|\mathbf{w}\|^2$ is minimized when $\mathbf{w}_\perp = 0$, since $\|\mathbf{w}_\Psi + \mathbf{w}_\perp\|^2 = \|\mathbf{w}_\Psi\|^2 + \|\mathbf{w}_\perp\|^2 \geq \|\mathbf{w}_\Psi\|^2$, where the first equality follows from orthogonality. **(+1pt)** Thus the optimal $\mathbf{w}$ lies in the row space of Ψ.

Part b)

$$\begin{aligned}
\mathcal{J}(\mathbf{w}) &= \frac{1}{2N}\|t - \Psi\mathbf{w}\|^2 + \frac{\lambda}{2}\|\mathbf{w}\|^2 \\
&= \frac{1}{2N}(t - \Psi\mathbf{w})^T(t - \Psi\mathbf{w}) + \frac{\lambda}{2}\mathbf{w}^T\mathbf{w} \\
&= \frac{1}{2}\alpha(\frac{1}{N}\mathbf{K}^T\mathbf{K} + \lambda\mathbf{K})\alpha - \frac{1}{N}t^T\mathbf{K}\alpha + \frac{1}{2N}t^T t. \text{ \textbf{(+2pt)}}
\end{aligned}$$

Plug in the minimum of this quadratic and simplify: $\alpha = (\mathbf{K} + \lambda N\mathbf{I})^{-1}t$. (The assumption that $\mathbf{K}$ is positive definite allows us to cancel the $\mathbf{K}$s in this step.) **(+1pt)**

2 Compositional Kernels

Here we use superscripts to denote coordinates of the feature vectors.

Part a)

$\Psi_S(\boldsymbol{x}) = [\Psi_1(\boldsymbol{x}); \Psi_2(\boldsymbol{x})] = [\Psi_1^{(1)}(\boldsymbol{x}); \ldots\Psi_1^{(m)}(\boldsymbol{x}); \Psi_2^{(1)}(\boldsymbol{x}); \ldots\Psi_2^{(n)}(\boldsymbol{x})]$. **(+1pt)**

Part b)

$\Psi_S(\boldsymbol{x}) = vec[\Psi_1(\boldsymbol{x}) \otimes \Psi_2(\boldsymbol{x})] = [\Psi_1^{(1)}(\boldsymbol{x})\Psi_2^{(1)}(\boldsymbol{x}); \ldots\Psi_1^{(m)}(\boldsymbol{x})\Psi_2^{(1)}(\boldsymbol{x}); \ldots\Psi_1^{(m)}(\boldsymbol{x})\Psi_2^{(n)}(\boldsymbol{x})]$.

- **(+1pt)** for writing the product of two kernels in terms of two feature maps
- **(+1pt)** for noting Ψ_S should contain products of elements in the two feature maps
- **(+1pt)** for taking the outer product of two feature maps and flatten.