Math 156 Project 1

Aatmun Baxi

Problem 1

Part 1

Let $k_a(\mathbf{x}_a, \mathbf{x}'_a) = \Phi_a(\mathbf{x}_a)^T \Phi_a(\mathbf{x}'_a)$, $k_b(\mathbf{x}_b, \mathbf{x}'_b) = \Phi_b(\mathbf{x}_b)^T \Phi_b(\mathbf{x}'_b)$ so that

$$k(\mathbf{x}, \mathbf{x}') = \Phi_a(\mathbf{x}_a)^T \Phi_a(\mathbf{x}'_a) + \Phi_b(\mathbf{x}_b)^T \Phi_b(\mathbf{x}'_b).$$

We claim that k is defined by a nonlinear feature space mapping

$$\psi(\mathbf{x}) = (\Phi_a(\mathbf{x}_a), \Phi_b(\mathbf{x}_b)).$$

where the subscripts a and b define the disjoint subsets of $\mathbf{x}$ as in the problem statement. This feature space mapping is valid as all components of $\mathbf{x}$ are present in the union of the sets $\mathbf{x}_a$, $\mathbf{x}_b$.

Evaluating an inner product as block matrices, we get that

$$\psi(\mathbf{x})^T \psi(\mathbf{x}') = (\Phi_a(\mathbf{x}_a), \Phi_b(\mathbf{x}_b))^T (\Phi_a(\mathbf{x}'_a), \Phi_b(\mathbf{x}'_b)).$$

Note here that inner product is valid since we can transpose the block vector components individually to get valid inner products in the expansion of the above expression. i.e.

$$\psi(\mathbf{x})^T \psi(\mathbf{x}') = (\Phi_a(\mathbf{x}_a), \Phi_b(\mathbf{x}_b))^T (\Phi_a(\mathbf{x}'_a), \Phi_b(\mathbf{x}'_b)) = \Phi_a(\mathbf{x}_a)^T \Phi_a(\mathbf{x}'_a) + \Phi_b(\mathbf{x}_b)^T \Phi_b(\mathbf{x}'_b),$$

which is exactly $k(\mathbf{x}, \mathbf{x}')$.

Part 2

Let $\mathbf{x} = (x_1, x_2)$, $\mathbf{y} = (y_1, y_2)$, so

$$k(\mathbf{x}, \mathbf{y}) = (x_1 y_1 + x_2 y_2)^2 = x_1^2 y_1^2 + 2 x_1 x_2 y_1 y_2 + x_2^2 y_2^2.$$

From here it is clear that the mapping

$$\Phi(\mathbf{x}) = \left(x_1^2,\ \sqrt{2} x_1 x_2,\ x_2^2\right)$$

produces the expression for k when taking $\Phi(\mathbf{x})^T \Phi(\mathbf{y})$.

Problem 2

The log likelihood of the model is

$$F(\mathbf{w}) = \frac{-\beta}{2} \sum_{n=1}^{N} [y(x_n, \mathbf{w}) - t_n]^2 + \frac{N}{2} \ln \beta - \frac{N}{2} \ln 2\pi.$$

To maximize we differentiate F with respect to each component of $\mathbf{w}$ $w_1, \ldots, w_M$. To compute this we compute the general partial derivatives of y with respect to each of the components $w_1, \ldots, w_M$.

$$\frac{\partial y}{\partial w_i} = x^i \qquad i = 0, \ldots, M.$$

Therefore

$$\frac{\partial F}{\partial x} = -\beta \sum_{n=1}^{N} \left[y\left(x_n, \mathbf{w}\right) - t_n \right] x_n^i \qquad i = 1, \ldots, M.$$

Setting each of these partials equal to 0 we get

$$\begin{aligned}
0 &= -\beta \sum_{n=1}^{N} \left[y\left(x_n, \mathbf{w}\right) - t_n \right] x_n^i \qquad i = 1, \ldots, M \\
\sum_{n=1}^{N} t_n x_n^i &= \sum_{n=1}^{N} x_n^i y\left(x_n, \mathbf{w}\right) \qquad i = 1, \ldots, M \\
\sum_{n=1}^{N} t_n x_n^i &= \sum_{n=1}^{N} x_n^i \sum_{j=0}^{M} w_j x_n^j \qquad i = 1, \ldots, M \\
\sum_{n=1}^{N} t_n x_n^i &= \sum_{n=1}^{N} \sum_{j=0}^{M} w_j x_n^{i+j} \qquad i = 1, \ldots, M \\
\sum_{n=1}^{N} t_n x_n^i &= \sum_{j=0}^{M} \sum_{i=1}^{N} w_j x_n^{i+j} \qquad i = 1, \ldots, M \\
\iff T_i &= \sum_{j=0}^{M} A_{ij} w_j
\end{aligned}$$

Problem 3

Let $A \in M_n\left(\mathbf{R}\right)$. Suppose $v \in \mathrm{Im}\left(A\right)$ so there exists $w \in \mathbf{R}^n$ such that $v = Aw$. Then

$$A\left(A^T A\right)^{-1} A^T v = A\left(A^T A\right)^{-1} A^T A w = A \mathbf{1}_n w = Aw = v$$

so the claimed projection matrix leaves elements in the image of A alone. Now suppose $v \in \mathrm{Im}\left(A\right)^{\perp}$. By rank nullity theorem this means that $v \in \mathrm{Ker}\left(A^T\right)$. So

$$A\left(A^T A\right)^{-1} A^T w = 0,$$

which is consistent with the behavior of orthogonal projections of vectors onto orthogonal spaces.

Problem 4

We differentiate in $\mathbf{w}$ and set equal to $\mathbf{0}$, solving for $\mathbf{w}$:

$$
\begin{aligned}
E_D(\mathbf{w}) &= \frac{1}{2}\sum_{n=1}^{N} r_n\{t_n - \mathbf{w}^T\phi(\mathbf{x}_n)\} : \\
E_D'(\mathbf{w}) &= \left(\sum_{n=1}^{N} r_n\{t_n - \mathbf{w}^T\phi(\mathbf{x}_n)\}\right)\left(-2\left(t_n - \mathbf{w}^T\phi(\mathbf{x}_n)\right)\phi(\mathbf{x}_n)\right) \\
\mathbf{0} &= \left(\sum_{n=1}^{N} r_n\{t_n - \mathbf{w}^T\phi(\mathbf{x}_n)\}\right)\left(-2\left(t_n - \mathbf{w}^T\phi(\mathbf{x}_n)\right)\phi(\mathbf{x}_n)\right) \\
\mathbf{0} &= \left(\sum_{n=1}^{N} r_n\{t_n - \mathbf{w}^T\phi(\mathbf{x}_n)\}\right)\left(-2t_n + 2\mathbf{w}^T\phi(\mathbf{x}_n)\right)\phi(\mathbf{x}_n) \\
\mathbf{0} &= \left(\sum_{n=1}^{N} r_n\{t_n - \mathbf{w}^T\phi(\mathbf{x}_n)\}\right)\phi(\mathbf{x}_n) \\
\mathbf{0}^T &= \left(\sum_{n=1}^{N} r_n\{t_n - \mathbf{w}^T\phi(\mathbf{x}_n)\}\right)\phi(\mathbf{x}_n)^T \\
\mathbf{0}^T &= \sum_{n=1}^{N} r_n t_n \phi^T(\mathbf{x}_n) - \sum_{n=1}^{N} r_n \mathbf{w}^T\phi(\mathbf{x}_n)\,\phi(\mathbf{x}_n)^T \\
\sum_{n=1}^{N} r_n t_n \phi(\mathbf{x}_n)^T &= \sum_{n=1}^{N} r_n \mathbf{w}^T\phi(\mathbf{x}_n)\,\phi(\mathbf{x}_n)^T \\
\sum_{n=1}^{N} r_n t_n \phi(\mathbf{x}_n) &= \sum_{n=1}^{N} r_n \phi(\mathbf{x}_n)\,\phi(\mathbf{x}_n)^T\,\mathbf{w} \\
\mathbf{w} &= \left(\sum_{n=1}^{N} r_n \phi(\mathbf{x}_n)\,\phi(\mathbf{x}_n)^T\right)^{-1}\left(\sum_{n=1}^{N} r_n t_n \phi(\mathbf{x}_n)\right)
\end{aligned}
$$

Problem 4

Final part

Using equation (1) we have

$$
\begin{aligned}
\left(w_n^{(k+1)}\right)^T \mathbf{x}_n &= \left(\left(\mathbf{w}^{(k)}\right)^T + \frac{\left(t_n - \left(\mathbf{w}^{(k)}\right)^T\right)\mathbf{x}_n}{\|\mathbf{x}_n\|^2}\mathbf{x}_n^T\right)\mathbf{x}_n \\
\left(w_n^{(k+1)}\right)^T \mathbf{x}_n &= \left(\mathbf{w}^{(k)}\right)^T \mathbf{x}_n + \frac{\left(t_n - \left(\mathbf{w}^{(k)}\right)^T\mathbf{x}_n\right)}{\|\mathbf{x}_n\|^2}\mathbf{x}_n^T\mathbf{x}_n \\
\left(w_n^{(k+1)}\right)^T \mathbf{x}_n &= \left(\mathbf{w}^{(k)}\right)^T \mathbf{x}_n + \left(t_n - \left(\mathbf{w}^{(k)}\right)^T\mathbf{x}_n\right) \\
\left(w_n^{(k+1)}\right)^T \mathbf{x}_n &= t_n
\end{aligned}
$$