HOMEWORK 2

VIKRAM GUPTA

1. Prob 2b.

Given that:

$$\mathcal{L}(\mathbf{w}, S) = \frac{1}{2}\sum_{n=1}^{N} c_n(y_n - \mathbf{w}^t\mathbf{x}_n)^2 + \frac{1}{2}\lambda\mathbf{w}^t\mathbf{w} \tag{1}$$

To convert (1) to matrix form, each c_n should be multiplied to individual $e_n = (y_n - \mathbf{w}^t\mathbf{x}_n)^2$ elements, and then their sum should be computed.

Let $C = [c_1, c_2, \ldots, c_N]^T$, and, for $\mathcal{L}$ to be a scalar quantity as expected, the matrix form of the equation should have a $N \times N$ matrix in place of C. The elements c_n and e_n should still be multiplied individually. Hence the C should be used as a diagonal matrix $\mathbf{C_d}$ where

$$\mathbf{C_d} = \begin{bmatrix} c_1 & 0 & 0 & \cdots \\ 0 & c_2 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ 0 & 0 & \cdots & c_n \end{bmatrix} \tag{2}$$

The matrix form of $\mathcal{L}$ can be written as:

$$\begin{aligned} \mathcal{L} &= \frac{1}{2}[(\mathbf{Y} - \mathbf{XW})^{\mathbf{T}}\mathbf{C_d}(\mathbf{Y} - \mathbf{XW})] + \frac{1}{2}\lambda\mathbf{W^TW} \\ &= \frac{1}{2}[\mathbf{Y^TC_dY} - \mathbf{Y^TC_dXW} - (\mathbf{XW})^{\mathbf{T}}\mathbf{C_dY} + \mathbf{W^TX^TC_dXW}] + \frac{1}{2}\lambda\mathbf{W^TW} \\ &= \frac{1}{2}[\mathbf{Y^TC_dY} - \mathbf{2W^TX^TC_dY} + \mathbf{W^TX^TC_dXW} + \lambda\mathbf{W^TW}] \end{aligned}$$

To find the normal equation we differentiate $\mathcal{L}$

$$\frac{d\mathcal{L}}{d\mathbf{W}} = 0$$

$$\frac{1}{2}[\phi - \mathbf{2X^TC_dY} + \mathbf{2X^TC_dXW} + 2\lambda\mathbf{W}] = 0$$

Hence,

$$\mathbf{X^TC_dY} = \mathbf{X^TC_dXW} + \lambda\mathbf{W}$$

And the normal equation is:

$$\mathbf{W} = (\mathbf{X^TC_dX} + \lambda\mathbf{I})^{-1}\mathbf{X^TC_dY}$$

where $\mathbf{I}$ is identity matrix of size $D \times D$, and, D is the number of dimensions in the data.