ROB 313 Assignment 2

Q1. $\hat{f}(\underline{x}, \underline{w}) = w_0 + \sum_{j=1}^{M-1} w_j \phi_j(\underline{x})$

$$\underset{\underline{w} \in \mathbb{R}^M}{\arg\min} \left(\sum_{i=1}^{N} \left(y^{(i)} - w_0 - \sum_{j=1}^{M-1} w_j \phi_j(\underline{x}^{(i)}) \right)^2 + \sum_{i=1}^{M} \sum_{j=1}^{M} \Gamma_{ij} w_{i-1} w_{j-1} \right)$$

rearrange to

$$\underset{\underline{w} \in \mathbb{R}^M}{\arg\min} \left(\sum_{i=1}^{N} \left(-\left(w_0 + \sum_{j=1}^{M-1} w_j \phi_j(\underline{x}^{(i)}) + y^{(i)} \right) \right)^2 + \sum_{i=1}^{M} \sum_{j=1}^{M} \Gamma_{ij} w_{i-1} w_{j-1} \right)$$

define $\underline{w} = \{w_0, w_1, \ldots, w_{M-1}\}^T \in \mathbb{R}^M$

and
$$\underline{\Phi} = \begin{bmatrix} \phi_0(\underline{x}^{(1)}) & \phi_1(\underline{x}^{(1)}) & \cdots & \phi_{M-1}(\underline{x}^{(1)}) \\ \phi_0(\underline{x}^{(2)}) & \phi_1(\underline{x}^{(2)}) & \cdots & \phi_{M-1}(\underline{x}^{(2)}) \\ \vdots & \vdots & & \vdots \\ \phi_0(\underline{x}^{(N)}) & \phi_1(\underline{x}^{(N)}) & \cdots & \phi_{M-1}(\underline{x}^{(N)}) \end{bmatrix} \in \mathbb{R}^{N \times M}$$

can rewrite
$$\sum_{i=1}^{N} \left(-\left(w_0 + \sum_{j=1}^{M-1} w_j \phi_j(\underline{x}^{(i)}) + y^{(i)} \right) \right)^2$$

as
$$(\underline{y} - \underline{\Phi}\underline{w})^T (\underline{I}) (\underline{y} - \underline{\Phi}\underline{w})$$

and rewrite
$$\sum_{i=1}^{M} \sum_{j=1}^{M} \Gamma_{ij} w_{i-1} w_{j-1}$$

as
$$\underline{w}^T \underline{\Gamma} \underline{w}$$

giving loss expression to be minimized as

$$\mathcal{L}(\underline{w}) = \left((\underline{y} - \underline{\Phi}\underline{w})^T (\underline{I}) (\underline{y} - \underline{\Phi}\underline{w}) + \underline{w}^T \underline{\Gamma} \underline{w} \right)$$

Q2. differentiating gets

$$\nabla_{\mathbf{w}} \mathcal{L} = 2 \mathbf{\Phi}^T (\mathbf{I}) \mathbf{\Phi} \mathbf{w} - 2 \mathbf{\Phi}^T (\mathbf{I}) \mathbf{y} + (\Gamma + \Gamma^T) \mathbf{w}$$

setting equal to zero gives

$$2 \mathbf{\Phi}^T (\mathbf{I}) \mathbf{\Phi} \mathbf{w} + 2 \mathbf{\Gamma} \mathbf{w} = 2 \mathbf{\Phi}^T (\mathbf{I}) \mathbf{y}$$

$$(2 \mathbf{\Phi}^T (\mathbf{I}) \mathbf{\Phi} + 2 \mathbf{\Gamma}) \mathbf{w} = 2 \mathbf{\Phi}^T \mathbf{y}$$

~~which is~~

$$\mathbf{w} = (2 \mathbf{\Phi}^T (\mathbf{I}) \mathbf{\Phi} + 2 \mathbf{\Gamma})^{-1} (2 \mathbf{\Phi}^T (\mathbf{I}) \mathbf{y})$$

which is an expression for the weights

Q2. Put $\hat{f}(\mathbf{x}, \boldsymbol{\alpha}) = \sum_{i=1}^{N} \alpha_i \, k(\mathbf{x}, \mathbf{x}^{(i)})$ into a matrix form along with the rest of the objective function to get

$$(\mathbf{K}(\mathbf{x}) \boldsymbol{\alpha} - \mathbf{y})^T (\mathbf{K}(\mathbf{x}) \boldsymbol{\alpha} - \mathbf{y}) + \lambda \boldsymbol{\alpha}^T \boldsymbol{\alpha}$$

Taking gradient and setting to zero gives

$$\nabla_{\alpha} = 2 \mathbf{K}(\mathbf{x})^T \mathbf{K}(\mathbf{x}) \boldsymbol{\alpha} - 2 \mathbf{K}^T(\mathbf{x}) \mathbf{y} + 2 \lambda \boldsymbol{\alpha} = 0$$

$$\underbrace{K(x)^T K(x)}_{\mathbf{K} \text{ (gram matrix)}} \alpha + \lambda \alpha = K^T(x) \mathbf{y}$$

giving

$$(\mathbf{K} + \lambda \mathbf{I}) \boldsymbol{\alpha} = K^T(x) \mathbf{y}$$

$$\boxed{\boldsymbol{\alpha} = \mathbf{K}^T(\mathbf{x}) \, (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{y}}$$

This the same as derived in class for weight vector which would be set to $\mathbf{w} = K(x) \boldsymbol{\alpha}$ giving the known $\boldsymbol{\alpha} = (\mathbf{K} + \lambda \mathbf{I})^{-1} \mathbf{y}$