594 HW2

Casey Sobecks

23 February 2021

1 Problem 2: Feature Mapping and Linear Regression

1.1 Question

Consider a data set $\mathcal{D} = \{x^{(n)}, y^{(n)}\}_{n=1}^{N}$ consisting of N scalar inputs $x^{(n)}$ and $y^{(n)}$. For a model consisting of a linear combination of M basis functions $\phi_j(x), j = 0, M-1$:

$$F(x, \mathbf{w}) = \mathbf{w}^T \phi(x)$$

with weights $\mathbf{w} = (w_0, ..., w_{M-1})^T$, show that the squared error cost function

$$\mathcal{C} = \frac{1}{2N} \sum_{n=1}^{N} \|F^{(n)}(x^{(n)}, \mathbf{w}) - y^{(n)}\|^2$$

is minimized for parameters:

$$\mathbf{w}^* = (\mathbf{\Phi}^T \mathbf{\Phi})^{-1} \mathbf{\Phi}^T \mathbf{y}$$

where $\mathbf{\Phi}$ is the feature matrix:

$$\mathbf{\Phi} = \begin{pmatrix} \phi_0(x^{(1)}) & \phi_1(x^{(1)}) & \dots & \phi_{M-1}(x^{(1)}) \\ \phi_0(x^{(2)}) & \phi_1(x^{(2)}) & \dots & \phi_{M-1}(x^{(2)}) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0(x^{(N)}) & \phi_1(x^{(N)}) & \dots & \phi_{M-1}(x^{(N)}) \end{pmatrix}$$

1.2 Solution

This problem is started by taking the derivative of the cost function.

$$\begin{aligned}
\frac{\partial C}{\partial w_k} &= \frac{1}{2N}\sum_n 2(\sum_j w_j\phi_j - y^{(n)})\frac{\partial}{\partial w_k}(\sum_i w_i\phi_i) \\
&= \frac{1}{N}\sum_n(\sum_j w_j\phi_j - y^{(n)})\sum_i(\phi_i\frac{\partial w_i}{\partial w_k}) \\
&= \frac{1}{N}\sum_n(\sum_j w_j\phi_j - y^{(n)})\sum_i \phi_i\delta_{ik} \\
&= \frac{1}{N}\sum_n(\sum_j w_j\phi_j - y^{(n)})\phi_k
\end{aligned}$$

With this, the optimal value of w can be found. To do this, the derivative just found will be set to zero. The optimal value will also be denoted as w^*. Solving is as follows,

$$0 = \frac{1}{N}\sum_n(\sum_j w_j\phi_j - y^{(n)})\phi_k$$

This can be rewritten as

$$\begin{aligned}
\sum_n\sum_j w_j^*\phi_j\phi_k &= \sum_n y^{(n)}\phi_k \\
\sum_n\sum_j w_j^*\phi_j^{(n)}\phi_k^{(n)} &= \sum_n y^{(n)}\phi_k^{(n)}
\end{aligned}$$

However, $\phi_j^{(n)} = \Phi_{nj}$. Using this identity,

$$\begin{aligned}
\sum_j w_j^*\sum_n \Phi_{nj}\Phi_{nk} &= \sum_n y^{(n)}\Phi_{nj} \\
\sum_j w_j^*\sum_n [\Phi^T]_{nj}\Phi_{nk} &= \sum_n y^{(n)}\Phi_{nj} \\
\sum_j w_j^*[\Phi^T\Phi]_{jk} &= \sum_n y^{(n)}\Phi_{nj} \\
[w^*\Phi^T\Phi]_j &= [\Phi^T y]_j \\
w^*\Phi^T\Phi &= \Phi^T y
\end{aligned}$$

With this equation, the problem is solved with some simple algebra.

$$w^* = (\Phi^T\Phi)^{-1}\Phi^T y$$