LAB 3 THEORY

NATHANIEL GUNTER

1. Problem 1:

1.1. **Problem Statement.** Derive a formula for which the coefficients $a^0, a^1, a^2, ..., a^m$,for which the linear function $y(x^1, x^2, ..., x^m) = a^0 + a^1x^1 + a^2x^2 + ... + a^mx^m$ minimizes
$E_2(a^0, a^1, a^2, ..., a^m) = \sum_{i=1}^N (y_i - (a^0 + a^1x_i^1 + a^2x_i^2 + ... + a^mx_i^m))^2$.

1.2. **Derivation.** Consider $y(x^1, x^2, ..., x^m)$ as a sum of $m+1$ functions, each of the form $\phi_i = x^i$ defining $\phi_0 = 1$. Then we have $y(\mathbf{x}, \mathbf{a}) = \mathbf{a}^T\phi(\mathbf{x})$. Now we must find the parameters $\mathbf{a}^T$.

We are given the cost function E_2, so to find a minimum, we set the gradient of the cost function to zero, that is $\nabla_{\mathbf{a}}E_2(\mathbf{a}) = 0$. For this cost function, this solves to $0 = \sum_{i=1}^N y_i\phi(\mathbf{x}_i)^T - \mathbf{a}^T(\sum_{i=1}^N \phi(\mathbf{x}_i)\phi(\mathbf{x}_i)^T)$.

This is then solved for the coefficients $\mathbf{a}$ and translated into matrix notation, giving

$$\mathbf{a} = (\mathbf{\Phi}^T\mathbf{\Phi})^{-1}\mathbf{\Phi}^T\mathbf{y}. \tag{1}$$

Where $\mathbf{y} = (y_1, ..., y_N)^T$ and $\mathbf{\Phi}$ is the matrix of predictor values where each row is given by $\Phi(i,:) = (1, x_i^1, ..., x_i^m)$.

In practice, this is simpler to solve in the form

$$A\mathbf{a} = \mathbf{\Phi}^T\mathbf{y}, \tag{2}$$

where $A = \mathbf{\Phi}^T\mathbf{\Phi}$.

Because of the structure of $\mathbf{\Phi}$, this results in

$$(A)_{ij} = \sum_{l=1}^N x_l^i x_l^j. \tag{3}$$

Because superscripts here are *not* exponents, this is not x_l^{i+j}. Similarly, the right hand side solves to be

$$b_i = \sum_{l=1}^N y_l x_l^i. \tag{4}$$

This pair of equations is what must be solved to find the coefficients $\mathbf{a}$.

Date: May 06, 2022.