Homework 3

CpE 4903 Neural Networks and Machine Learning

Multiple Linear Regression & Linear Algebra

Samiya Zaidi (szaidi4)

September 10, 2023

1 Linear Algebra Concepts

1. $\boxed{\text{10 points}}$ **Matrix Multiplication.**
   Given 2 matrices, A and B, please manually calculate $C = A \cdot B$.

$$A = \begin{bmatrix} 2 & 1 \\ 3 & 4 \\ 1 & 2 \end{bmatrix}, \; B = \begin{bmatrix} 3 & 2 & 1 & 4 \\ 1 & 0 & 2 & 1 \end{bmatrix}$$

Solution: We know that the resultant matrix, C, will be a 3×4 matrix.

$$C = \begin{bmatrix} (2\times3)+(1\times1) & (2\times2)+(1\times0) & (2\times1)+(1\times2) & (2\times4)+(1\times1) \\ (3\times3)+(4\times1) & (3\times2)+(4\times0) & (3\times1)+(4\times2) & (3\times4)+(4\times1) \\ (1\times3)+(2\times1) & (1\times2)+(2\times0) & (1\times1)+(2\times2) & (1\times4)+(2\times1) \end{bmatrix}$$

$$C = \begin{bmatrix} 6+1 & 4+0 & 2+2 & 8+1 \\ 9+4 & 6+0 & 3+8 & 12+4 \\ 3+2 & 2+0 & 1+4 & 4+2 \end{bmatrix}$$

$$\therefore C = \begin{bmatrix} 7 & 4 & 4 & 9 \\ 13 & 6 & 11 & 16 \\ 5 & 2 & 5 & 6 \end{bmatrix}$$

2. $\boxed{\text{10 points}}$ **Matrix Transpose**.

Given matrix $A = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}$, please find $B = A^T \cdot A$. Show that B is symmetric.

Solution: Take the transpose of A^T:

$$A^T = \begin{bmatrix} a & d \\ b & e \\ c & f \end{bmatrix}$$

Now find B.

$$B = \begin{bmatrix} a^2 + d^2 & ab + de & ac + df \\ ba + ed & b^2 + e^2 & bc + ef \\ ca + fd & cb + ef & c^2 + f^2 \end{bmatrix}$$

Since multiplication for numbers is commutative, B can be written as:

$$B = \begin{bmatrix} a^2 + d^2 & ab + de & ac + df \\ ab + de & b^2 + e^2 & bc + ef \\ ac + df & bc + ef & c^2 + f^2 \end{bmatrix}$$

B is a symmetric matrix because the entries across the main diagonal are exactly the same. This happened because a matrix was being multiplied by it's transpose, which always produces a symmetric matrix (theorem).

3. $\boxed{\text{10 points}}$ **Solving Linear Equations**
Given the following system of three linear equations:

$$\begin{cases} 2x + 3y - z = 7 \\ x–2y + 2z = 3 \\ 3x + y - 4z = 4 \end{cases}$$

Rewrite the system as its equivalent matrix equation, $A \cdot x = b$. Define A, x, and b explicitly.

Solution: The given system of linear equations can also be written in the following form:

$$\begin{bmatrix} 2 & 3 & -1 \\ 1 & -2 & 2 \\ 3 & 1 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 7 \\ 3 \\ 4 \end{bmatrix}$$

In this case, A is the coefficient matrix, x has the unknowns and b is the resultant matrix:

$$A = \begin{bmatrix} 2 & 3 & -1 \\ 1 & -2 & 2 \\ 3 & 1 & -4 \end{bmatrix},\ x = \begin{bmatrix} x \\ y \\ z \end{bmatrix},\ b = \begin{bmatrix} 7 \\ 3 \\ 4 \end{bmatrix}$$

2 Multiple Linear Regression and Linear Algebra

1. $\boxed{\text{10 points}}$ **Creating the Feature Matrix**

   Based on the dataset, please create the Feature Matrix, $\vec{X}$. First, write down the the definition of $\vec{X}$ then populate the matrix with the values from the dataset. Also, what is the shape (number of rows and number of columns) of this matrix?

   Solution: The feature matrix is a representation of a large dataset in a matrix form. The entries of the matrix represent the values of a certain parameter. The row represents the values of all parameters for the same sample, whereas, the columns represent the values of a particular parameter/feature for all the samples.
   In this case, the feature matrix is:

   $$\vec{X} = \begin{bmatrix} 5 & 21 & 45 & 1 \\ 3 & 14 & 40 & 2 \\ 2 & 8.5 & 30 & 1 \end{bmatrix}$$

   To consider, the bias term, a_0, we need to add one more column, containing all ones. This is not a real parameter, but it's only included to make the multiplication easier.

   $$\vec{X} = \begin{bmatrix} 1 & 5 & 21 & 45 & 1 \\ 1 & 3 & 14 & 40 & 2 \\ 1 & 2 & 8.5 & 30 & 1 \end{bmatrix}$$

   With this additional column, the shape of the matrix would be 3×5. Because it has 3 rows and 5 columns.

2. $\boxed{\text{15 points}}$ **Predicting Housing Prices.**

   Assume the weight vector $A = (a_0, a_1, a_2, a_3, a_4) = (780, 18, 40, -25, -50)$, perform the following:

   Write down the vector equation for predicting the target value $\hat{y}$ of the first training example using the weight vector and the features, $x^{(1)} = (x_1^{(1)}, x_2^{(1)}, x_3^{(1)}, x_4^{(1)}) = (5, 21, 45, 1)$.

   Solution: The vector equation can be written in as the matrix multiplication between the transpose of A and the feature vector, x.

   $$\hat{y} = x \cdot A^T$$

   If we substitute the actual matrices, it would be:

   $$\hat{y} = \begin{bmatrix} 1 & 5 & 21 & 45 & 1 \end{bmatrix} \cdot \begin{bmatrix} 780 \\ 18 \\ 40 \\ -25 \\ -50 \end{bmatrix}$$

Note: I concatenated a column of one with the feature matrix to include a value for the bias term, a_0. Without this, the matrix multiplication will be an invalid operation.

3. $\boxed{\text{15 points}}$ **Calculating the Cost Function**.

Given the weight vector, $A = (780, 18, 40, 25, 50)$, and the dataset provided earlier, please calculate the cost function, $J(A)$.

Write down the matrix formula for the cost function, in the context of multiple linear regression.

Solution: The matrix formula of $J(A)$ can be written with the help of the vector notation.

$$\frac{1}{2\times 3}\left(\begin{bmatrix}1 & 5 & 21 & 45 & 1\\ 1 & 3 & 14 & 40 & 2\\ 1 & 2 & 8.5 & 30 & 1\end{bmatrix}\cdot\begin{bmatrix}780\\18\\40\\-25\\-50\end{bmatrix}-\begin{bmatrix}460\\232\\178\end{bmatrix}\right)^T\cdot\left(\begin{bmatrix}1 & 5 & 21 & 45 & 1\\ 1 & 3 & 14 & 40 & 2\\ 1 & 2 & 8.5 & 30 & 1\end{bmatrix}\cdot\begin{bmatrix}780\\18\\40\\-25\\-50\end{bmatrix}-\begin{bmatrix}460\\232\\178\end{bmatrix}\right)$$

4. $\boxed{\text{14 points}}$ **Finding the Optimum Weight Vector: Normal Equation**.
Write down the matrix formula with the numerical values.

Solution: The closed form solution for A_{opt} is

$$A_{opt} = (\vec{X}^T\cdot\vec{X})^{-1}\cdot\vec{X}^T\cdot Y$$

Write in the form of matrices;

$$A_{opt} = \left(\begin{bmatrix}1 & 1 & 1\\ 5 & 3 & 2\\ 21 & 14 & 8.5\\ 45 & 40 & 30\\ 1 & 2 & 1\end{bmatrix}\cdot\begin{bmatrix}1 & 5 & 21 & 45 & 1\\ 1 & 3 & 14 & 40 & 2\\ 1 & 2 & 8.5 & 30 & 1\end{bmatrix}\right)^{-1}\cdot\begin{bmatrix}1 & 1 & 1\\ 5 & 3 & 2\\ 21 & 14 & 8.5\\ 45 & 40 & 30\\ 1 & 2 & 1\end{bmatrix}\cdot\begin{bmatrix}460\\232\\178\end{bmatrix}$$

5. $\boxed{\text{15 points}}$ **Calculating the Gradient Vector**.
Given the weight vector, $A = (780, 18, 40, 25, 50)$, and the dataset provided earlier, please calculate the gradient vector of the cost function.

Write down the matrix expression for the gradient of the cost function.

Solution: The gradient can be calculated using the partial derivatives:

$$\frac{\partial J(A)}{\partial a_j} = \frac{1}{m}\sum_{i=1}^{m}(X^{(i)} \cdot A^T - y^{(i)})x_j^{(i)},\ \forall j$$

The vector equation can be written as:

$$\frac{\partial J(A)}{\partial A} = \frac{1}{m}\begin{bmatrix} \frac{\partial J(A)}{\partial a_0} \\ \\ \frac{\partial J(A)}{\partial a_1} \\ \\ \frac{\partial J(A)}{\partial a_2} \\ \\ \frac{\partial J(A)}{\partial a_3} \\ \\ \frac{\partial J(A)}{\partial a_4} \end{bmatrix} = \frac{1}{m}(X^T) \cdot (X \cdot A^T - Y)$$