COMPSCI 4ML3

Introduction to Machine Learning

Winter 2021

Assignment One

Tahseen Ahmed

ahmedt26

McMaster University

Revised: February 2nd, 2020

Hello, and welcome to my submission for assignment one. Below you will find all of my solutions to the requested questions from the assignment instructions.

1. **Question One**

   We find the best line $\hat{y} = ax + b$ that fits $f(x) = 2x^2 + x^3$. We know x is uniformly distributed on $[0, 1]$. We know since $n \to \infty$, we can convert the summation to an integral.

   $$\mathsf{OLS} : \Delta = \sum_{i=1}^{n} (a.x^i + b - y^i)^2$$
   $$\therefore \Delta' = \int_0^1 (ax + b - (2x^2 + x^3))^2 dx$$

   Now that we've derived Δ', we use our partial derivatives to get a and b by setting $\frac{\partial \Delta'}{\partial a} = 0$ and $\frac{\partial \Delta'}{\partial b} = 0$.

   $\frac{\partial}{\partial a}$**:**

   $$\frac{\partial}{\partial a} \int_0^1 (ax + b - (2x^2 + x^3))^2 dx = \int_0^1 2x(ax + b - (2x^2 + x^3))dx$$
   $$\frac{2}{3}a + b = \frac{7}{5}$$

   $\frac{\partial}{\partial b}$**:**

   $$\frac{\partial}{\partial b} \int_0^1 (ax + b - (2x^2 + x^3))^2 dx = \int_0^1 2(ax + b - (2x^2 + x^3))dx$$
   $$a + 2b = \frac{11}{6}$$

   We now solve for a and b. Note $a = \frac{11}{6} - 2b$. We will substitute this into the first equation we calculated to get b.

   $$\frac{2}{3}(\frac{11}{6} - 2b) + b = \frac{7}{5}$$
   $$\frac{11}{9} - \frac{4}{3}b + b = \frac{7}{5}$$
   $$-\frac{1}{3}b = \frac{8}{45}$$
   $$b = -\frac{8}{15} \approx -0.53333$$

   Now we compute a:

   $$a + 2b = \frac{11}{6}$$
   $$a - \frac{16}{15} = \frac{11}{6}$$
   $$a = \frac{29}{10}$$

   We have therefore computed $\hat{y} = \frac{29}{10}x + \frac{-8}{15}$. Using a modified ols.ipynb helps confirm these values.

2. **Question Two**

Below is the target function:

$$f(x) = 2 + 2x_1 + 3x_2 + 4x_3 + 3x_1x_2 - 5x_2x_3 + 2x_1^2x_3^2$$

We compute our hyperplane to be:

$$\hat{y} = -0.0016978x_1 + 0.00381558x_2 + 0.00367768x_3 + 6.954947950300704$$

Here is the code. I have not included solve_ols and run_ols since they're exactly the same from the given file. I used the SciKit Learn module as a secondary source to confirm my solutions. SciKit Learn and the given ols code give slightly different outputs. I used the SciKit code to compute the regressor values since the output looked cleaner to me. I used this site for the code:

```
# COMPSCI 4ML3 Q2 - Second Try
# Tahseen Ahmed
# Saturday, February 7th, 2020
# ==============================
import matplotlib.pyplot as plt
import numpy as np
from sklearn.linear_model import LinearRegression # Use this for MLR

# generate n data points based on a
# combination of sinosuidal and polynomial functions
def generate_data(n):
    X = np.random.rand(n, 3)
    Y = 2 + (2*X[0]) + (3*X[1]) + (4*X[2]) + (3 *X[0] * X[1])
          - (5 * X[1] * X[2]) + (2 * X[0]**2 * X[2]**2)
    # Adding noise
    Y = Y + 0.1*np.random.randn(n,1)
    return X, Y

# Number of training and test points.
n_train = 25000
n_test  = 25000

# This will be used later for regularization. For now it is set to 0.
alpha = 0

# Generating train and test data.
X_train, Y_train = generate_data(n_train)
X_test, Y_test = generate_data(n_test)

# Three-dimensional hyperplane (y = ax1 + bx2 + cx3 + d)
# Data Augmentation:
X_augmented_train = np.concatenate((X_train,np.ones((n_train, 1))), axis=1)
```

```
X_augmented_test = np.concatenate((X_test,np.ones((n_test, 1))), axis=1)

# Now we run OLS on the augmented data.
run_ols(X_augmented_train, Y_train, X_augmented_test,
             Y_test, alpha, X_train, X_test, "y = ax1 + bx2 + cx3 + d")

print('\n')
print("==============================================")
print("     SciKit Learn Linear Model")
print("==============================================")
print('\n')
regressor = LinearRegression()
regressor.fit(X_train,Y_train)

# display coefficients, not sure about intercept,
# so I take the mean of the matrix.
print(regressor.coef_, regressor.intercept_.mean())
```

Note the estimated hyperplane is indicating there is no real trend in the data (regressors all near 0). The most notable aspect is the hyperplane is situated at $y \approx 7$.

The regressors values gained vary considerably each run, but they all are near 0.

3. **Question Three**

We would like to fit $\hat{y} = ax$ on the curve $y = x^2 + x$. We will minimize this objective function:

$$\sum_i \left[(\frac{y^i}{\hat{y}^i})^2 + (\frac{\hat{y}^i}{y^i})^2 \right]$$

We assume x is uniform on $[0, 1]$ so we can use integrals instead of the summations.

$$\Delta' = \int_0^1 \left[(\frac{x^2 + x}{ax})^2 + (\frac{ax}{x^2 + x})^2 \right]$$

$$= \int_0^1 \left[\frac{(x+1)^2}{a^2} + \frac{a^2}{(x+1)^2} \right]$$

$$\frac{\partial \Delta'}{\partial a} = \int_0^1 \frac{2a}{(x+1)^2} - \frac{2(x+1)^2}{a^3}$$

$$\int_0^1 \frac{2a}{(x+1)^2} - \frac{2(x+1)^2}{a^3} = a - \frac{14}{3a^3}$$

Set the above equal to 0:

$$0 = a - \frac{14}{3a^3}$$

$$a = \frac{14}{3a^3}$$

$$1 = \frac{14}{3a^4}$$

s

$$a^4 = \frac{14}{3}$$

$$a = 1.46977784$$

Computing the value using our code gets someting around 1.7486, which is rougly a 0.27 difference in slope.

4. **Question Four**

   Recall that X is an $n \times d$ matrix. In this case $n < d$.

   A straightforward reason why $X^T X$ won't be invertible is because the new matrix will not be linearly dependent since you are multiplying elements in the matrix with another, hence some rows/columns will lose independence if they had any. $(X^T X)$ is $d \times d$, and those new columns are not independent. The matrix is no longer of full rank. Such a matrix has a determinant of 0, and we know that this results in the matrix to not be invertible, which is also considered a degenerate case.

   Typically when this occurs, the problem can't actually be solved since there is an infinite amount of solutions. We can add a bias term to figure out a solution.

5. **Question Five**

Here are some tips gathered from the MS Teams Discussion and the question:

- X is an $n \times n$ matrix, so square.
- $\mathsf{Rank}(X) = n$, which means X is of full-rank.
- X is linearly independent.
- $X^T X$ is invertible because of the above statement.

We write down some useful formulas gathered to solve this proof, their lettering (a and b) will be used in the two-column proof.:

a. $\Delta = XW^{LS} - Y$
b. $W^{LS} = (X^T X)^{-1} X^T Y$

The hint from the assignments tells us to use $X^T \Delta$ and the theorem of linear independence. We will keep this in mind. We will also use theorems and properties given by the complimentary reading found in the discussions for Q4. namely from section 3.7, on the inverse of matrices.

Proof. We begin to prove that $\| XW^{LS} - Y \|_2^2 = 0$.

$\| XW^{LS} - Y \|_2^2$	Starting definition
$\| X((X^T X)^{-1} X^T Y) - Y \|_2^2$	Theorem b
$\| XX^{-1} X^{-T} X^T Y - Y \|_2^2$	$(AB)^{-1} = B^{-1}A^{-1}$ applied to $(X^T X)^{-1}$
$\| I_{n\times n} X^{-T} X^T Y - Y \|_2^2$	$XX^{-1} = I_{n\times n}$ since X is invertible
$\| I_{n\times n} I_{n\times n} Y - Y \|_2^2$	$X^{-T} X^T = I_{n\times n}$ since X^T and X are invertible
$\| Y - Y \|_2^2$	$I_{n\times n} M_{n\times n} = I_{n\times n}$, applied twice, $M \neq 0_{n\times n}$
$\| 0_{n\times n} \|_2^2$	$M_{n\times n} - M_{n\times n} = 0_{n\times n}$
0	Norm of a zero vector is zero

By actively using the knowledge of X being invertible, full-rank etc, we have proved that $\| XW^{LS} - Y \|_2^2 = 0$.

□