M20 - Introduction to Computer Programming with MATLAB
Instructor: Prof. Enrique López Droguett, Ph.D.
Teacher Assistants: M. Fidansoy, G. San Martín, M. Pishahang, V. Vela.
Fall 2023 – UCLA
Student: *Alex Lie*
UCLA ID: *905901892*

HOMEWORK 6

Task 1: Linear Regression

Introduction

I am using MATLAB to create linear regressions to model United States wildfire data from 1985 to 2022. The number of wildfires, the burned acres, and the firefighting costs will be graphed against time. I will report the formula for the linear regressions, as well as the coefficient of determinations. I will make predictions about wildfire statistics in 2025 and I will model firefighting cost as a function of burned acres and run a linear regression. Lastly, I will repeat this using data only from 1985 to 2020.

Model and Theory

N/A

Methodology

I first cleared the workspace and command window. I then imported the data from the Excel spreadsheet using the readtable() function and declaring variables. I then ran three distinct linear regressions using the linReg() function. This function takes in an independent variable and a dependent variable as parameters. The year is the independent variable for all three regressions, and the dependent variable changes. The number of wildfires, the burned acres, and the firefighting cost are the dependent variables we model against time. The linReg() function solves the system of the two linear normal equations. The output of the linReg() function includes the coefficients of the linear regression, the coefficient of determination, the standard deviation of the dependent variable, and the standard error of the dependent variable around the regression line. I then outputted the formula for the three linear regressions and the coefficient of determination values using fprintf(). I then created three plots, modeling the actual data points and the corresponding fitted regression lines. I created x and y values using linspace() and my linear regression formulas, then made the plots by using subplot(). I added titles, axes labels and axes, limits. Using my linear regressions, I calculated the predicted number of wildfires, the burned acres, and the firefighting costs in the United States in 2025. After, I ran a linear regression using acres as the independent variable and cost as the dependent variable. I then outputted the formula for the linear regression as well as the corresponding coefficient of determination. I then plotted the data and the regression line, and estimated the firefighting cost in a year when 7.5 million acres are consumed by wildfires using the linear regression model. Afterwards, I removed the last two rows of the original data in order to recreate the linear regression excluding the last two years, since the firefighting costs exhibited an unexpected increase. I declared a new vector, from the original vector, but the first parameter is “1:end-2”, and the second parameter is “:”). I then ran linReg(), outputted the formula for the linear regression and the coefficient of determination, created my plot, and estimated the firefighting cost when 7.5 million acres are consumed by wildfires using the formula. Lastly, I added comments to make my code easier to understand for users.

Calculations and results

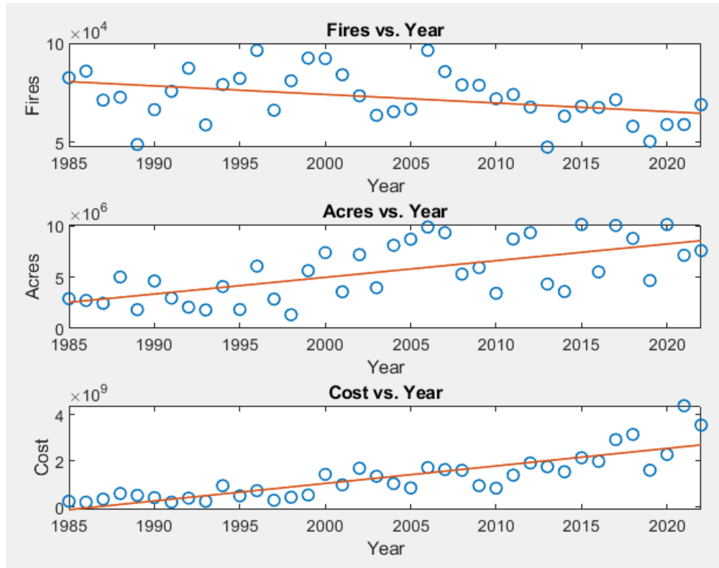
See next page.

M20 - Introduction to Computer Programming with MATLAB
Instructor: Prof. Enrique López Droguett, Ph.D.
Teacher Assistants: M. Fidansoy, G. San Martín, M. Pishahang, V. Vela.
Fall 2023 – UCLA
Student: *Alex Lie*
UCLA ID: *905901892*

1)

```
Fires vs. Year
y = -432.167 * x + 938484.590
r-squared = 0.145294

Acres vs. Year
y = 162405.678 * x + -319834372.022
r-squared = 0.416146

Cost vs. Year
y = 76021676.332 * x + -151026995057.884
r-squared = 0.718982
```

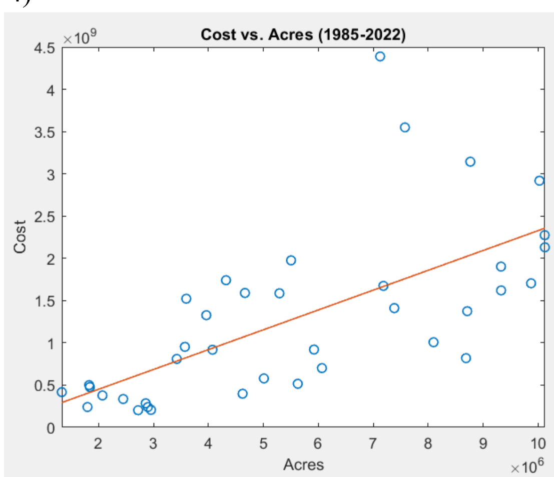
2)



3)

```
fires_2025 =

   6.3346e+04


acres_2025 =

   9.0371e+06


cost_2025 =

   2.9169e+09
```

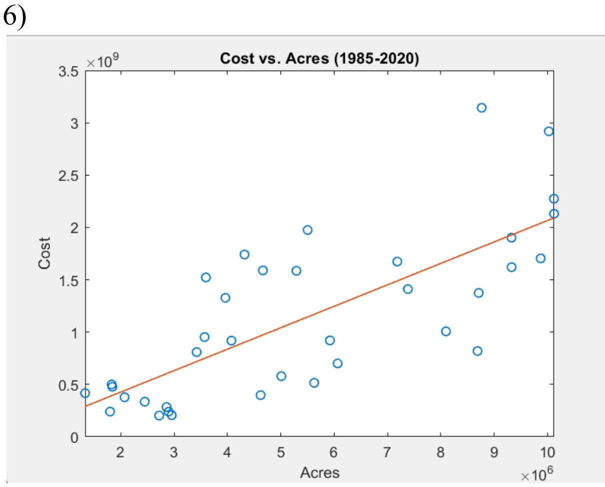
4)



5)

```
Cost vs. Acres (1985-2022)
y =  234.593 * x + -18481280.515
r-squared = 0.433941


cost1_prediction =

   1.7576e+10
```

6)



7)

```
Cost vs. Acres (1985-2020)
y =  204.823 * x + 17906770.748
r-squared = 0.560869


cost2_prediction =

   1.5380e+10
```

Student: *Alex Lie*
UCLA ID: *905901892*

Discussions and Conclusions

By looking at the coefficient of determinations from the first three linear regressions (1), the wildfire data is best modeled as cost as a function of year because the r-squared value is the highest (0.718982). The other two coefficients of determinations (0.145294 for fires vs. year) and (0.416146 for acres vs. year) are lower in value and are not as strong to form a linear relationship. By looking at the three plots (2), we can see the data points are close to the regression line for cost vs. year, but not as close for acres vs. year and fires vs. year. The cost vs. year and acres vs. year regression line slopes upwards, while the fires vs. year regression line slopes downwards. In the United States in 2025, the predicted number of fires is 63346, the predicted number of acres burned is 9.0371 million, and the predicted firefighting cost is 2916.9 million (3). The linear regression plots using 1985-2022 data and 1985-2020 data (4 and 6) are very similar, but two data points are removed in the 1985-2020 plot. The 1985-2020 data slopes slightly less upwards compared to the 1985-2022 plot. The estimated firefighting cost if 7.5 million acres are burned using the 1985-2022 model is 17.576 billion and the estimated firefighting cost using the 1985-2020 model is 15.380 billion (5 and 7). The difference is about 2.196 billion. The coefficient of determination in the 1985-2022 model is 0.433941. The coefficient of determination in the 1985-2020 model is 0.560869, which is larger than the coefficient of determination in the other model by 0.126928. Removing the data from the last two years makes the linear regression fit our data better, since there is an increase in the coefficient of determination.

Task 2: Multiple Linear Regression

Introduction

Continuing from Task 1, I am using MATLAB to create a multiple linear regression to estimate firefighting costs as a function of the number of fires and the burned acres. I will report the formula for the multiple linear regression and the coefficients. I will then repeat this, but using data only from 1985 to 2020. I will then predict the firefighting cost in a year using the two multiple linear regression formulas. I will also make 3D graphs of both of the regression planes to compare the two models.

Model and Theory

N/A

Methodology

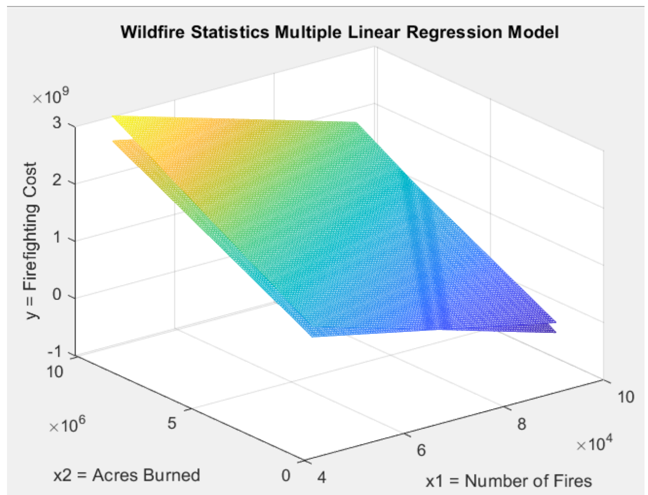
I first cleared the workspace and command window. I then imported the data from the Excel spreadsheet using the readtable() function and declaring variables. I then clarified what my x1, x2, and y variables are. x1 represents the number of fires, x2 represents the burned acres, and y represents the firefighting cost. I then solved the system of three linear equations to find the coefficients to the multiple linear regression. I first declared the basis function matrix Z with its corresponding values. I then declared the column vector Y. Lastly, I solved for the coefficient matrix C by setting it equal to the inverse of Z multiplied by Y. I then outputted the multiple linear regression formula and the coefficients used in that multiple linear regression. Next, I then removed the last two rows from the original data in order to recreate the multiple linear regression excluding the last two years, since the firefighting costs exhibited an unexpected increase. I redeclared Z and Y, and resolved for C. I then outputted my new formula and coefficients. Afterwards, I estimated the firefighting costs in a year if 70k fires consume 7.5 million acres with both models using the multiple linear regression formulas. I then graphed the regression planes. I created values for x1 (number of fires) and x2 (acres burned) using linspace(), used meshgrid(), and created values for y (cost) by using the multiple linear regression formulas. I then used mesh() to model both regression planes in the same figure. I added labels and a title. Lastly, I added comments to make my code easier to understand for other users.

Calculations and results

1)

```
x1 = number of fires
x2 = burned acres
y = firefighting cost

Multiple Linear Regression (1985-2022)
y =  235.908 * x2 + -25773.503 * x1 + 1846366035.842
a0 = 1846366035.842
a1 = -25773.503
a2 =  235.908

Multiple Linear Regression (1985-2020)
y =  207.780 * x2 + -18371.476 * x1 + 1345101179.492
a0 = 1345101179.492
a1 = -18371.476
a2 =  207.780

Cost Prediction (1985-2022)
   1.8115e+09

Cost Prediction (1985-2020)
   1.6174e+09
```

2)



The firefighting cost prediction if 70k fires consume 7.5 million acres using the 1985-2022 model is 181.15 million. The firefighting cost prediction if 70k fires consume 7.5 million acres using the 1985-2022 model is 161.74 million. The graph shows the two different regression planes using the 1895-2022 model and the 1895-2020 model.

Discussions and Conclusions

By looking at figure 1, the two different predictions of the firefighting cost differ by 200 million. The coefficients for the linear regression are different as well. I believe that the 1985-2020 model is more accurate than the 1985-2022 model because the 1985-2020 model excluded the outliers with the higher than normal firefighting cost in the most recent years.

Analyzing figure 2, both regression planes look similar to each other and are close to each other.