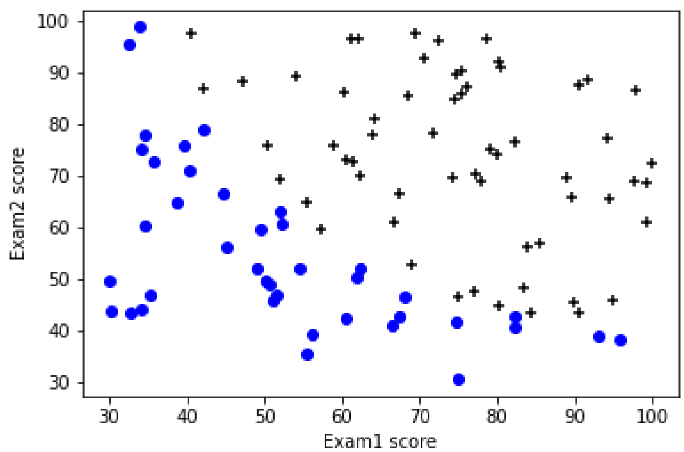
Rayan Hassan
4511021

ECE 1395 – Homework 3 report

1) a) The size of matrix X and Y are shown below. X has 100 rows and 3 columns (Exam1 scores, Exam2 scores and x0), and Y has 100 rows (for each training sample) and 1 column (decision)

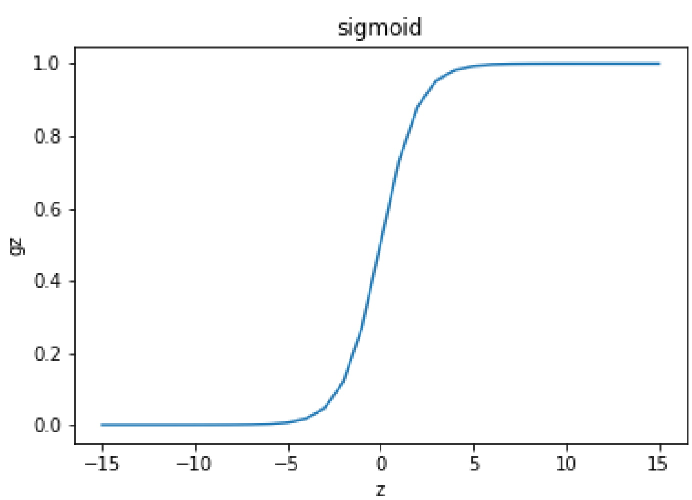
```
In [3]: runfile('C:/Users/RAYAN/OneDrive/Desktop/
ps3.py', wdir='C:/Users/RAYAN/OneDrive/Desktop')
Reloaded modules: sigmoid, costFunction, gradFunction,
normalEqn
Size of matrix X is:  (100, 3)
Size of matrix Y is:  (100, 1)

In [4]:
```

b) This is the scatter plot of training data (saved as ps3-1-b.png)



c) Data divided randomly into training set and test set using test_train_split (view code)

d) The following shows the gz versus z plot for the sigmoid function (ps3-1-c.png)



Rayan Hassan
4511021

e) Here is the cost value obtained from function costFunction for the toy set

```
Cost J for the toy set when theta=[0,0,0] is:  0.6931471805599453

In [13]:
```

And this is the gradient of the cost obtained using gradFunction

```
Gradient of the cost of the toy set is:  [[ 0.   ]
 [ 0.   ]
 [-0.75]]
```

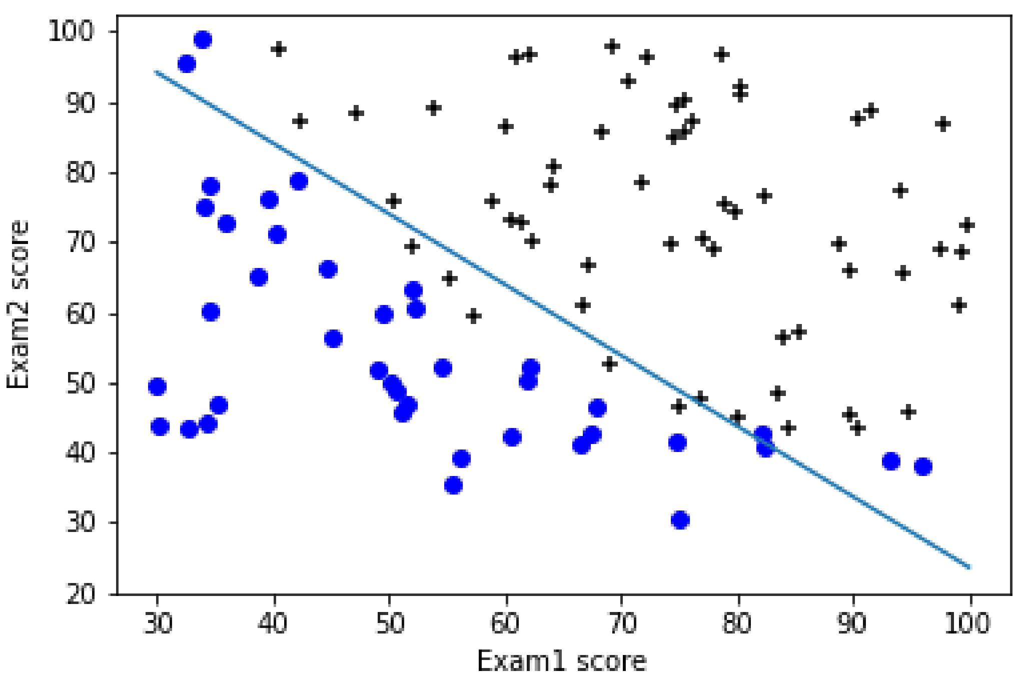
f) The following screenshot shows output of the min_bfgs function as long as the optimal parameter theta.

```
Optimization terminated successfully.
         Current function value: 0.216343
         Iterations: 21
         Function evaluations: 29
         Gradient evaluations: 29
The optimal parameter theta is:  [-23.87964054   0.19345199   0.19185897]
```

We can see that the optimal parameters theta = [-23.8796, 0.1934, 0.1918]

And the value of the cost function at convergence Is 0.216343

g) The decision boundary can be seen next (saved as ps3-1-f.png)



Rayan Hassan
4511021

h) This is the code to calculate the accuracy, as well as the resulting output

```
#----------------------Accuracy of logistic regression----------------#
theta = np.array([[op1[0]],[op1[1]],[op1[2]]])

h_test= np.dot(X_test,theta)

h_test = h_test>0  #if True (1) -> admitted, else not admitted

correctly_classified = 0
for i in range(len(h_test)):
    if int(h_test[i][0]) == Y_test[i][0]:
        correctly_classified += 1

accuracy = correctly_classified/len(Y_test)
print("Accuracy of logistic regression is: ",accuracy)
```

```
Accuracy of logistic regression is:  1.0
```

i) The admission probability for student who got 80 on test1 and 50 on test 2 is shown below

```
Admission probability is:  76.66457328124595 %
```

The decision should be "admitted"

j) BONUS. Please see the pictures below

Rayan Hassan
4511021

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y\log(h_\theta(x)) + (1-y)\log(1-h_\theta(x))\right]$$

Applying the chain rule:

$$\frac{\partial J(\theta)}{\partial\theta_j} = -\frac{1}{m}\sum_{i=1}^{m}\left[y\,\frac{1}{h_\theta(x)}\,\frac{\partial h_\theta(x)}{\partial\theta_j} + (1-y)\,\frac{1}{1-h_\theta(x)}\,\frac{\partial(1-h_\theta(x))}{\partial\theta_j}\right]$$

* Now let's find $\frac{\partial h_\theta(x)}{\partial\theta_j}$

we know $h_\theta(x) = \frac{1}{1+e^{-\theta^T x}}$

Let's find the derivative of $\sigma(x) = \frac{1}{1+e^{-x}}$

$$\frac{d\,\sigma(x)}{dx} = \frac{d}{dx}\left(1+e^{-x}\right)^{-1}$$

$$= -1\left(1+e^{-x}\right)^{-2}\left(-e^{-x}\right)$$

$$= \frac{-e^{-x}}{-(1+e^{-x})^2} = \frac{e^{-x}}{(1+e^{-x})^2}$$

$$= \frac{1}{1+e^{-x}}\,\frac{e^{-x}}{1+e^{-x}} = \frac{1}{1+e^{-x}}\cdot\frac{(1+e^{-x})-1}{1+e^{-x}}$$

$$= \frac{1}{1+e^{-x}}\cdot\left[\cancel{\frac{1+e^{-x}}{1+e^{-x}}} - \frac{1}{1+e^{-x}}\right]$$

$$= \frac{1}{1+e^{-x}}\left[1-\frac{1}{1+e^{-x}}\right] = \boxed{\sigma(x)\left(1-\sigma(x)\right)}$$

So now
$$\frac{\partial h_\theta(x)}{\partial\theta_j} = \underbrace{h_\theta(x)\left(1-h_\theta(x)\right)}\,\frac{\partial\theta^T x}{\partial\theta_j}$$

following the pattern of the derivative of sigmoid function.

Similarly
$$\frac{\partial(1-h_\theta(x))}{d\theta_j} = -h_\theta(x)\left(1-h_\theta(x)\right)\frac{\partial\theta^T x}{\partial\theta_j}$$

and we know $\frac{\partial\theta^T x}{\partial\theta_j} = x_j^i$

So we can write:

Rayan Hassan
4511021

$$\frac{\partial J(\theta)}{\partial \theta_j} = -\frac{1}{m}\sum_{i=1}^{m}\Big[y \frac{1}{\cancel{h_\theta(x)}} \cancel{h_\theta(x)}\,(1-h_\theta(x))\, x_j^i$$

$$+ (1-y)\frac{1}{\cancel{1-h_\theta(x)}}(-h_\theta(x))\,\cancel{(1-h_\theta(x))}\, x_j^i\Big]$$

$$= \frac{-1}{m}\sum_{i=1}^{m}\left[y\,(1-h_\theta(x))\,x_j^i + (1-y)(-h_\theta(x))\,x_j^i\right]$$

$$= \frac{-1}{m}\sum_{i=1}^{m}\left[y\,(1-h_\theta(x)) - (1-y)\,h_\theta(x)\right] x_j^i$$

$$= \frac{-1}{m}\sum_{i=1}^{m}\left[y - \cancel{y h_\theta(x)} - h_\theta(x) + \cancel{y h_\theta(x)}\right] x^i$$

$$= \frac{-1}{m}\sum_{i=1}^{m}\left[y - h_\theta(x)\right] x_j^i$$

$$= \boxed{\frac{1}{m}\sum_{i=1}^{m}(h_\theta(x) - y)\, x_j^i}$$

final answer

2) a) The learned model parameter is shown below:

```
The learned model parameter is:  [[ 2.19256629e+05]
 [-7.75886157e+02]
 [ 1.06170508e+01]]
```

b) The data and the model are plotted on the following figure (saved as ps3-2-b.png)

Rayan Hassan
4511021