CS145 Howework 1

Important Note: HW1 is due on **11:59 PM PT, Oct 19 (Monday, Week 3)**. Please submit through GradeScope (you will receive an invite to Gradescope for CS145 Fall 2020.).

Print Out Your Name and UID

Name: Rithika Srinivasan, UID: 905125793

Before You Start

You need to first create HW1 conda environment by the given `cs145hw1.yml` file, which provides the name and necessary packages for this tasks. If you have `conda` properly installed, you may create, activate or deactivate by the following commands:

```
conda env create -f cs145hw1.yml
conda activate hw1
conda deactivate
```

OR

```
conda env create --name NAMEOFYOURCHOICE -f cs145hw1.yml
conda activate NAMEOFYOURCHOICE
conda deactivate
```

To view the list of your environments, use the following command:

```
conda env list
```

More useful information about managing environments can be found here.

You may also quickly review the usage of basic Python and Numpy package, if needed in coding for matrix operations.

In this notebook, you must not delete any code cells in this notebook. If you change any code outside the blocks that you are allowed to edit (between `STRART/END YOUR CODE HERE`), you need to highlight these changes. You may add some additional cells to help explain your results and observations.

In [5]:
```
import numpy as np
import pandas as pd
import sys
import random as rd
import matplotlib.pyplot as plt
%load_ext autoreload
%autoreload 2
```

```
The autoreload extension is already loaded. To reload it, use:
  %reload_ext autoreload
```

If you can successfully run the code above, there will be no problem for environment setting.

1. Linear regression

This workbook will walk you through a linear regression example.

In [6]:
```
from hw1code.linear_regression import LinearRegression

lm=LinearRegression()
lm.load_data('./data/linear-regression-train.csv','./data/linear-regression-test
# As a sanity check, we print out the size of the training data (1000, 100) and
print('Training data shape: ', lm.train_x.shape)
print('Training labels shape:', lm.train_y.shape)
```

```
Training data shape:  (1000, 100)
Training labels shape: (1000,)
```

1.1 Closed form solution

In this section, complete the `getBeta` function in `linear_regression.py` which use the close for solution of $\hat{\beta}$.

Train you model by using `lm.train('0')` function.

Print the training error and the testing error using `lm.predict` and `lm.compute_mse` given.

In [7]:
```
from hw1code.linear_regression import LinearRegression

lm=LinearRegression()
lm.load_data('./data/linear-regression-train.csv','./data/linear-regression-test
training_error= 0
testing_error= 0
#========================#
# STRART YOUR CODE HERE  #
#========================#
lm.normalize()
beta = lm.train('0')
predicted_train_y = lm.predict(lm.train_x, beta)
training_error = lm.compute_mse(predicted_train_y, lm.train_y)

predicted_test_y = lm.predict(lm.test_x, beta)
testing_error = lm.compute_mse(predicted_test_y, lm.test_y)

#========================#
#   END YOUR CODE HERE   #
#========================#
print('Training error is: ', training_error)
print('Testing error is: ', testing_error)
```

```
Learning Algorithm Type:  0
y shape
(1000,)
x shape
(1000, 101)
Beta:  [ 5.23000000e-01 -3.95099505e-02 -3.01401932e-02 -5.71438644e-02
 -1.72769796e-02 -4.13700127e-03 -5.86318630e-02 -6.89027284e-02
```

```
 -3.56331805e-02 -1.87845537e-02 -1.82888714e-02  5.29276130e-02
  2.53519018e-02 -4.15812928e-02 -3.30193382e-02  2.65867992e-03
  1.34068950e-02 -3.88013327e-02  4.11038867e-02 -2.32239983e-02
 -2.68494719e-02 -5.67582270e-02  2.85948574e-02 -5.22058491e-02
 -1.94232592e-02  4.61988692e-02 -3.87491283e-02  3.82055256e-02
  1.27021593e-02  5.82271850e-02 -4.20937718e-02 -8.05582038e-02
  5.50688227e-02 -2.88202457e-02 -1.94706479e-02  2.58596756e-03
  2.55048685e-02  1.39991237e-02 -3.38312079e-02 -1.80218433e-02
 -8.42135902e-03 -5.61252496e-02 -3.60939866e-02 -1.12787490e-04
 -4.02969672e-02 -1.20851201e-02 -1.41809480e-02  5.11770552e-03
  4.48842190e-02  1.42864924e-02  1.79066117e-02 -3.08841654e-02
 -3.67139837e-02  3.83560781e-02 -4.47435146e-02 -6.08180754e-02
  4.69774181e-02 -5.86346690e-02  1.62361334e-02 -6.06942237e-02
 -3.38205570e-02 -4.24317897e-02 -5.46648364e-02 -2.89378305e-02
  5.33687506e-02 -3.17462303e-02  2.12826319e-03 -3.26837546e-02
  6.84819052e-03  1.25455103e-02 -4.09640271e-02  8.88512549e-03
  1.94883628e-02  6.04797247e-02 -4.23185183e-02 -4.76582979e-02
 -6.69833777e-02  5.66019062e-02  4.63178581e-03  4.13664903e-02
  7.10828556e-02  4.08986579e-02 -6.46605942e-02  3.05062530e-02
  6.11970818e-02 -6.13118531e-04  4.12093831e-02  8.04511196e-05
  3.21203863e-02  5.30651849e-02 -2.83935172e-02 -4.22856651e-02
  4.23271015e-02 -1.72635991e-03 -6.75124152e-02 -3.30151234e-02
 -2.14687553e-03 -6.00152621e-02  4.30059659e-02  6.79904935e-02
 -3.84367853e-03]
Training error is:  0.08693886675396784
Testing error is:  0.11017540281675804
```

1.2 Batch gradient descent

In this section, complete the `getBetaBatchGradient` function in `linear_regression.py` which compute the gradient of the objective fuction.

Train you model by using `lm.train('1')` function.

Print the training error and the testing error using `lm.predict` and `lm.compute_mse` given.

In [8]:

```python
lm=LinearRegression()
lm.load_data('./data/linear-regression-train.csv','./data/linear-regression-test
training_error= 0
testing_error= 0
#========================#
# STRART YOUR CODE HERE  #
#========================#
# lm.normalize()
beta2 = lm.train('1')
predicted_train_y = lm.predict(lm.train_x, beta2)
training_error = lm.compute_mse(predicted_train_y, lm.train_y)

predicted_test_y = lm.predict(lm.test_x, beta2)
testing_error = lm.compute_mse(predicted_test_y, lm.test_y)



#========================#
#   END YOUR CODE HERE   #
#========================#
print('Training accuracy is: ', training_error)
print('Testing accuracy is: ', testing_error)
```

```
Learning Algorithm Type:  1
Training accuracy is:  0.08694299599095222
```

```
Testing accuracy is:  0.11024614457982801
```

1.3 Stochastic gadient descent

In this section, complete the `getBetaStochasticGradient` function in `linear_regression.py`, which use an estimated gradient of the objective function.

Train you model by using `lm.train('2')` function.

Print the training error and the testing error using `lm.predict` and `lm.compute_mse` given.

In [9]:
```
lm=LinearRegression()
lm.load_data('./data/linear-regression-train.csv','./data/linear-regression-test
training_error= 0
testing_error= 0
#========================#
# STRART YOUR CODE HERE  #
#========================#
lm.normalize()
beta3 = lm.train('2')
print("beta shape: {}".format(beta3.shape))
predicted_train_y = lm.predict(lm.train_x, beta3)
training_error = lm.compute_mse(predicted_train_y, lm.train_y)

predicted_test_y = lm.predict(lm.test_x, beta3)
testing_error = lm.compute_mse(predicted_test_y, lm.test_y)
#========================#
#   END YOUR CODE HERE   #
#========================#
print('Training accuracy is: ', training_error)
print('Testing accuracy is: ', testing_error)
```

```
Learning Algorithm Type:  2
beta: [ 5.27260629e-01 -3.22490083e-02 -2.52778364e-02 -7.06036388e-02
 -2.82354510e-02  3.56668202e-02 -4.70891581e-02 -7.57249454e-02
 -2.35042789e-02  2.81274688e-03 -4.46334966e-03  6.95732170e-02
  2.43536507e-02 -5.23413242e-02 -1.89250970e-02  1.04886265e-02
 -3.16493883e-03 -3.63267560e-02  4.29911807e-02 -2.95306646e-02
 -1.49627856e-02 -7.07073352e-02  4.11758087e-02 -1.78392945e-02
 -2.45146168e-02  7.94358034e-02 -3.84546251e-02  4.05421998e-02
  3.51370715e-02  1.08159615e-01 -3.04196076e-02 -1.09225607e-01
  4.18097838e-02 -2.14100077e-02 -2.09306075e-02  2.01561313e-02
  2.31726355e-02  2.09414656e-02 -5.71102529e-02  3.43026875e-04
 -1.50662680e-02 -7.01256778e-02 -2.35008893e-02 -1.53134156e-02
 -3.74197261e-02 -1.51730935e-02  5.14151802e-04 -1.79295083e-02
  5.07170746e-02  7.32814765e-03  2.52680257e-02 -3.44471047e-02
 -4.41691822e-02  5.56941732e-02 -4.59119668e-02 -4.61664008e-02
  5.50444992e-02 -6.34538920e-02  4.02294047e-02 -2.33090825e-02
 -2.34403532e-02 -2.03412452e-02 -6.36716510e-02 -4.47915248e-02
  6.92972447e-02 -6.44205665e-02 -2.58217647e-02 -3.33776821e-02
  6.83578939e-03  2.79441471e-02 -6.38713402e-02  3.35247472e-03
  1.16311083e-02  5.02485442e-02 -3.93100840e-02 -6.56327722e-02
 -6.44397514e-02  4.40650187e-02 -1.03643931e-02  3.61635994e-02
  9.81052000e-02  2.70771613e-02 -6.81477490e-02  5.67842272e-02
  4.95537479e-02  7.87280014e-03  2.64716020e-02  3.97973853e-04
  3.75332601e-02  5.89671590e-02 -3.37741464e-02 -5.44337262e-02
  5.69887385e-02 -1.68717547e-02 -5.32183805e-02 -6.45138326e-02
 -3.32494904e-02 -7.15423468e-02  3.39220730e-02  7.38992551e-02
  1.20473436e-02]
Beta:  [ 5.27260629e-01 -3.22490083e-02 -2.52778364e-02 -7.06036388e-02
 -2.82354510e-02  3.56668202e-02 -4.70891581e-02 -7.57249454e-02
```

```
 -2.35042789e-02  2.81274688e-03 -4.46334966e-03  6.95732170e-02
  2.43536507e-02 -5.23413242e-02 -1.89250970e-02  1.04886265e-02
 -3.16493883e-03 -3.63267560e-02  4.29911807e-02 -2.95306646e-02
 -1.49627856e-02 -7.07073352e-02  4.11758087e-02 -1.78392945e-02
 -2.45146168e-02  7.94358034e-02 -3.84546251e-02  4.05421998e-02
  3.51370715e-02  1.08159615e-01 -3.04196076e-02 -1.09225607e-01
  4.18097838e-02 -2.14100077e-02 -2.09306075e-02  2.01561313e-02
  2.31726355e-02  2.09414656e-02 -5.71102529e-02  3.43026875e-04
 -1.50662680e-02 -7.01256778e-02 -2.35008893e-02 -1.53134156e-02
 -3.74197261e-02 -1.51730935e-02  5.14151802e-04 -1.79295083e-02
  5.07170746e-02  7.32814765e-03  2.52680257e-02 -3.44471047e-02
 -4.41691822e-02  5.56941732e-02 -4.59119668e-02 -4.61664008e-02
  5.50444992e-02 -6.34538920e-02  4.02294047e-02 -2.33090825e-02
 -2.34403532e-02 -2.03412452e-02 -6.36716510e-02 -4.47915248e-02
  6.92972447e-02 -6.44205665e-02 -2.58217647e-02 -3.33776821e-02
  6.83578939e-03  2.79441471e-02 -6.38713402e-02  3.35247472e-03
  1.16311083e-02  5.02485442e-02 -3.93100840e-02 -6.56327722e-02
 -6.44397514e-02  4.40650187e-02 -1.03643931e-02  3.61635994e-02
  9.81052000e-02  2.70771613e-02 -6.81477490e-02  5.67842272e-02
  4.95537479e-02  7.87280014e-03  2.64716020e-02  3.97973853e-04
  3.75332601e-02  5.89671590e-02 -3.37741464e-02 -5.44337262e-02
  5.69887385e-02 -1.68717547e-02 -5.32183805e-02 -6.45138326e-02
 -3.32494904e-02 -7.15423468e-02  3.39220730e-02  7.38992551e-02
  1.20473436e-02]
beta shape: (101,)
Training accuracy is:  0.11373741198992812
Testing accuracy is:  0.13756413236233234
```

Questions:

1. Compare the MSE on the testing dataset for each version. Are they the same? Why or why not?
2. Apply z-score normalization for eachh featrure and comment whether or not it affect the three algorithm.
3. Ridge regression is adding an L2 regularization term to the original objective function of mean squared error. The objective function become following:

$$J(\beta) = \frac{1}{2n}\sum_i \left(x_i^T\beta - y_i\right)^2 + \frac{\lambda}{2n}\sum_j \beta_j^2,$$

   where $\lambda \geq 0$, which is a hyper parameter that controls the trade off. Take the derivative of this provided objective function and derive the closed form solution for β.

Please type your answer here!

Your answer here:

1. The MSE for closed form solution is very similar to the descents, around 0.1, because they are essentially producing the same functions, but the batch solution's beta is more of an approximate of inversing the matrix, by calculating iteratively instead of all at once. The stochastic gradient descent example is based off of a random order of values with each time its function is called, so it changes every time it is called.
2. Adding z-score normalization does not affect the first, closed form feature, since the data values are not affected at each step. However, the z score normalization raises the MSE for the batch gradient, because it does not account for outliers that it does not map. Also, the

stochastic gradient example would not work without normalized data at all because the path the gradient follows would be impossible to discover.

3.

$$J(\beta) = \frac{1}{2n}\sum_i (x_i^T\beta - y_i)^2 + \underbrace{\frac{\lambda}{2n}\sum_j \beta_j^2}_{\text{L2 regularization term}}$$

in matrix form:

$$J(\beta) = \frac{(X\beta - y)^T(X\beta - y)}{2n} + \frac{\lambda}{2n}\beta^T\beta$$

$$= \frac{1}{2n}(\beta^T X^T - y^T)(X\beta - y) + \frac{\lambda}{2n}\beta^T\beta$$

$$= \frac{1}{2n}\left(\beta^T X^T X\beta - y^T X\beta - \beta^T X^T y + y^T y\right) + \frac{\lambda}{2n}\beta^T\beta$$

$$\frac{dJ}{d\beta} = \frac{1}{\cancel{2}n}\left(\cancel{2}X^T X\beta - \cancel{2}X^T y\right) + \frac{\cancel{2}\lambda}{\cancel{2}n}\beta = 0$$

$$= \frac{1}{n}\left(X^T X\beta - X^T y + \lambda\beta\right) = 0$$

$$= \frac{1}{n}\left((X^T X + \lambda)\beta - X^T y\right) = 0$$

$$= \frac{X^T X + \lambda}{n}\beta - \frac{X^T y}{n} = 0$$

$$= \frac{X^T X + \lambda}{n}\beta = \frac{X^T y}{n}$$

$$\boxed{\beta = (X^T X + \lambda)^{-1} X^T y}$$

2. Logistic regression

This workbook will walk you through a logistic regression example.

In [37]:

```
from hw1code.logistic_regression import LogisticRegression

lm=LogisticRegression()
```

```
lm.load_data('./data/logistic-regression-train.csv','./data/logistic-regression-
# As a sanity chech, we print out the size of the training data (1000, 5) and tr
print('Training data shape: ', lm.train_x.shape)
print('Training labels shape:', lm.train_y.shape)
```

```
Training data shape:  (1000, 5)
Training labels shape: (1000,)
```

2.1 Batch gradiend descent

In this section, complete the `getBeta_BatchGradient` in `logistic_regression.py`, which compute the gradient of the log likelihoood function.

Complete the `compute_avglogL` function in `logistic_regression.py` for sanity check.

Train you model by using `lm.train('0')` function.

And print the training and testing accuracy using `lm.predict` and `lm.compute_accuracy` given.

In [41]:
```
lm=LogisticRegression()
lm.load_data('./data/logistic-regression-train.csv','./data/logistic-regression-
training_accuracy= 0
testing_accuracy= 0
#========================#
# STRART YOUR CODE HERE  #
#========================#
lm.normalize()
beta4 = lm.train('0')

# predicted_train_y = lm.predict(lm.train_x, beta4)
# training_error = lm.compute_accuracy(predicted_train_y, lm.train_y)

# predicted_test_y = lm.predict(lm.test_x, beta4)
# testing_error = lm.compute_accuracy(predicted_test_y, lm.test_y)
#========================#
#   END YOUR CODE HERE   #
#========================#
print('Training accuracy is: ', training_accuracy)
print('Testing accuracy is: ', testing_accuracy)
```

```
average logL for iteration 0: -2.0960887433574817
/Users/rithika/Desktop/sad145/hw1/hw1code/logistic_regression.py:81: RuntimeWarn
ing: invalid value encountered in double_scalars
  stupid = exponent_beta_T / (1 + exponent_beta_T)
average logL for iteration 1000: nan
average logL for iteration 2000: nan
average logL for iteration 3000: nan
average logL for iteration 4000: nan
average logL for iteration 5000: nan
average logL for iteration 6000: nan
average logL for iteration 7000: nan
average logL for iteration 8000: nan
average logL for iteration 9000: nan
Training avgLogL:  nan
Training accuracy is:  0
Testing accuracy is:  0
```

2.2 Newton Raphhson

In this section, complete the `getBeta_Newton` in `logistic_regression.py` , which make use of both first and second derivative.

Train you model by using `lm.train('1')` function.

Print the training and testing accuracy using `lm.predict` and `lm.compute_accuracy` given.

In [ ]:
```
lm=LogisticRegression()
lm.load_data('./data/logistic-regression-train.csv','./data/logistic-regression-
training_accuracy= 0
testing_accuracy= 0
#========================#
# STRART YOUR CODE HERE  #
#========================#

#========================#
#   END YOUR CODE HERE   #
#========================#
print('Training accuracy is: ', training_accuracy)
print('Testing accuracy is: ', testing_accuracy)
```

Questions:

1. Compare the accuracy on the testing dataset for each version. Are they the same? Why or why not?
2. Regularization. Similar to linear regression, an regularization term could be added to logistic regression. The objective function becomes following:

$$J(\beta) = -\frac{1}{n}\sum_i \left(y_i x_i^T \beta - \log\left(1 + \exp\{x_i^T \beta\}\right)\right) + \lambda \sum_j \beta_j^2,$$

where $\lambda \geq 0$, which is a hyper parameter that controls the trade off. Take the derivative $\frac{\partial J(\beta)}{\partial \beta_j}$ of this provided objective function and provide the batch gradient descent update.

Your answer here:

Please type your answer here!

2.3 Visualize the decision boundary on a toy dataset

In this subsection, you will use the same implementation for another small dataset with each datapoint x with only two features (x_1, x_2) to visualize the decision boundary of logistic regression model.

In [ ]:
```
from hw1code.logistic_regression import LogisticRegression

lm=LogisticRegression(verbose = False)
lm.load_data('./data/logistic-regression-toy.csv','./data/logistic-regression-to
# As a sanity chech, we print out the size of the training data (99,2) and train
print('Training data shape: ', lm.train_x.shape)
print('Training labels shape:', lm.train_y.shape)
```

In the following block, you can apply the same implementation of logistic regression model (either in 2.1 or 2.2) to the toy dataset. Print out the $\hat{\beta}$ after training and accuracy on the train set.

In [ ]:
```
training_accuracy= 0
#========================#
# STRART YOUR CODE HERE  #
#========================#

#========================#
#   END YOUR CODE HERE   #
#========================#
print('Training accuracy is: ', training_accuracy)
```

Next, we try to plot the decision boundary of your learned logistic regression classifier. Generally, a decision boundary is the region of a space in which the output label of a classifier is ambiguous. That is, in the given toy data, given a datapoint $x = (x_1, x_2)$ on the decision boundary, the logistic regression classifier cannot decide whether $y = 0$ or $y = 1$.

Question

Is the decision boundary for logistic regression linear? Why or why not?

Your answer here:

Please type your answer here!

Draw the decision boundary in the following cell. Note that the code to plot the raw data points are given. You may need `plt.plot` function (see here).

In [ ]:
```
# scatter plot the raw data
df = pd.concat([lm.train_x, lm.train_y], axis=1)
groups = df.groupby("y")
for name, group in groups:
    plt.plot(group["x1"], group["x2"], marker="o", linestyle="", label=name)

# plot the decision boundary on top of the scattered points
#========================#
# STRART YOUR CODE HERE  #
#========================#

#========================#
#   END YOUR CODE HERE   #
#========================#
plt.show()
```

End of Homework 1 :)

After you've finished the homework, please print out the entire `ipynb` notebook and two `py` files into one PDF file. Make sure you include the output of code cells and answers for questions. Prepare submit it to GradeScope.