In [1]:

```
### Put your NAME and EID here: Ayush Srivastava (as79973)
```

# Problem Set 03b

Make sure you have the following packages installed for Python3:

- scikit-learn
- numpy
- matplotlib

In [2]:

```
# imports needed
import numpy as np
import matplotlib.pyplot as plt
import sklearn
from sklearn import datasets

from sklearn.linear_model import LogisticRegression
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
import sklearn.model_selection

# setting seed, DON'T modify
np.random.seed(10)


from pylab import rcParams
rcParams['figure.figsize'] = (10, 7)
```

## Problem 1: Classifying Boston Dataset

## Part A.

In this homework, you will be getting hands on experience on training your own linear classifiers. To do this will be working with the [Boston dataset (https://www.cs.toronto.edu/~delve/data/boston/bostonDetail.html)](https://www.cs.toronto.edu/~delve/data/boston/bostonDetail.html).

Luckily, scikit-learn has this dataset available to use. This part will have you setup their dataset to be used for classification:

- First, download the boston dataset from **scikit-learn**.
  - They should have a module that allows you to download it directly: [load_boston (https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_boston.html)](https://scikit-learn.org/stable/modules/generated/sklearn.datasets.load_boston.html)
  - After calling **load_boston()**, it should return a dictionary-like object that contains information.
  - The data itself is found in the **.data** item in the dictionary.
- Notice that .data has shape **(506,13)**. It does not have any given labels, but we will create our own synthetic labels.
  - Extract the first column from this matrix (i.e. index 0).
  - Calculate the median of this column.
- Now we will assign our own **y** based on the following:
  - Assign a class of **0** to each sample (row) if the first column is **less than** the above median.
  - Assign **1** otherwise (i.e. it has a value **greater than or equal to** the median).
- Finally, we will define our dataset:
  - **X**: all data samples using every feature **EXCEPT** for the first column.
    - This should have shape (506,12)
  - **y**: created from the previous step.

Useful modules:

```
- sklearn.datasets.load_boston
- np.median
```

In [3]:

```
data = datasets.load_boston().data
print("Data Shape :",data.shape)
col0 = data[:,0]
median=np.median(col0)
print("Median: ",median)
y = (col0>=median)*1
X= data[:,1:13]
X.shape
```

```
Data Shape : (506, 13)
Median:  0.25651
```

Out[3]:

```
(506, 12)
```

## Part B.

Now we will train two separate classifiers for this task -- Logistic Regression and Linear Discriminant Analysis, using scikit-learn.

- As usual, divide X,y into separate training/testing sets.
  - Use the **first 400 samples** as **Xtrain, ytrain**
  - Use the **next 106 samples** as **Xtest, ytest**
- Now create a logistic regression model using scikit-learn, and train it on the **training set**.
  - This will be similar to the previous homework, but using their logistic regression module.
  - Only use the default parameters (i.e. don't pass any extra parameters to the scikit-learn module)
  - Measure the following **prediction error** i.e. (1-accuracy) on the **test set**.
- Now create an LDA model using scikit-learn, and repeat the steps above.
- For ease of grading, please create a **bar chart** that shows the error of both the LogisticRegression and LDA models.
  - Make sure to label each bar accordingly and include a legend.

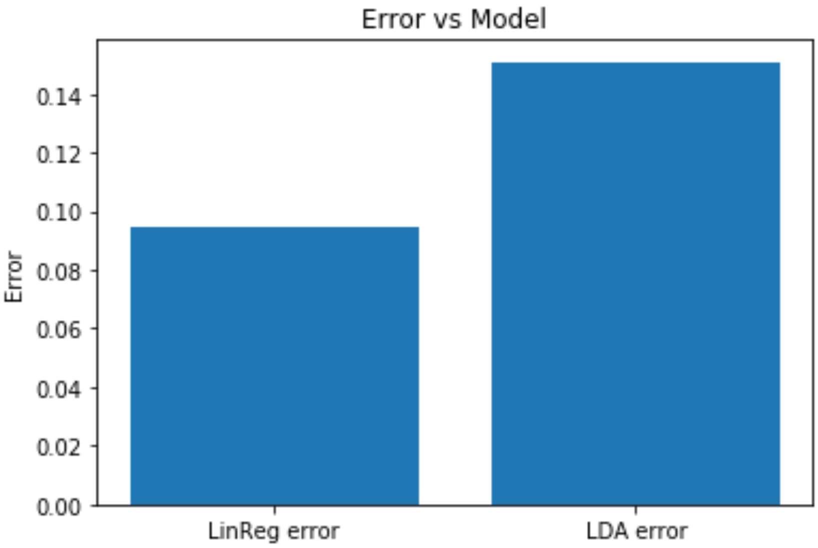
Useful modules:

```
- sklearn.linear_model.LogisticRegression
- sklearn.discriminant_analysis.LinearDiscriminantAnalysis
- plt.bar
- plt.legend
```

In [4]:

```
Xtrain = X[:400]
Xtest = X[400:]
Ytrain = y[:400]
Ytest = y[400:]

logReg = LogisticRegression().fit(Xtrain,Ytrain)
LDA = LinearDiscriminantAnalysis().fit(Xtrain,Ytrain)
logScore = logReg.score(Xtest,Ytest)
ldaScore = LDA.score(Xtest,Ytest)
print("error  Logreg ",1-logScore)
print("error LDA= ",1-ldaScore)
objects = ('LinReg error','LDA error')
y_pos = np.arange(len(objects))
performance = [1-logScore,1-ldaScore]

plt.bar(y_pos, performance, align='center')
plt.xticks(y_pos, objects)
plt.ylabel('Error')
plt.title('Error vs Model')

plt.show()
```

```
error  Logreg  0.09433962264150941
error LDA=  0.15094339622641506

/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
```



## Part C.

For this last section, we will be running an experiment that measures the effect of the # of training samples vs. the test set error.

**Note:** In this section, make sure to keep your **test set constant**, and only do the following operations on the **train set**.

Let our possible $n = [100, 200, 300, 400]$. For **BOTH** LogisticRegression and LDA, and for each $n$, do the following:

- Take a random sample from the training set of size $n$. Assign these to new variables, do not overwrite the original training set.
- Now train your model using this smaller random sample.
- Collect the **test prediction error**.
- Now, repeat the above steps **10 times**. Find the mean for these prediction error samples, and the standard deviation.

Once you have collected the mean prediction error and standard deviations for each $n$ and for each model you will now visualize:

- Create a bar plot with the **x-axis** as $n$ and the **y-axis** mean test prediction error.
- This bar plot should contain the following:
  - Two bars for each $n$, one for LogisticRegression, the other LDA. Make them different colors.
  - In addition add in **error bars** based on your computed standard deviations.
  - Include a legend, and axis titles.

Useful modules:

```
- sklearn.linear_model.LogisticRegression
- sklearn.discriminant_analysis.LinearDiscriminantAnalysis
- np.std
- np.mean
- plt.bar
- plt.legend
```

In [5]:

```
for n in [100,200,300,400]:
    log_error = []
    lda_error = []
    for j in range(0,100,10):
        X_train, X_test, Y_train, Y_test = sklearn.model_selection.train_test_split(X, y, test_size=n,random_state=j)
        clfLDA = LinearDiscriminantAnalysis(n_components=1)
        clfLDA.fit_transform(X_train, Y_train)
        temp = 1-clfLDA.score(X_test, Y_test)
        lda_error.append(temp)
        clfLog = LogisticRegression().fit(X_train, Y_train)
        temp = 1-clfLog.score(X_test, Y_test)
        log_error.append(temp)


    log_error_array = np.asarray(log_error)
    lda_error_array = np.asarray(lda_error)
    performance  = [lda_error_array.mean(),log_error_array.mean()]
    performance_deviation = [ lda_error_array.std(),log_error_array.std()]
    objects = ( 'LDA','Logistic')
    y_pos = np.arange(len(objects))
    plt.bar(y_pos + ((n-100)/50), performance, yerr = performance_deviation,align='center', alpha=0.5, label="Samples: %d" % n)
    plt.xticks(y_pos, objects)
    plt.title('Error vs Model vs Sample size')
    plt.ylabel('Prediction Failure rate')

plt.legend()
plt.show()
```

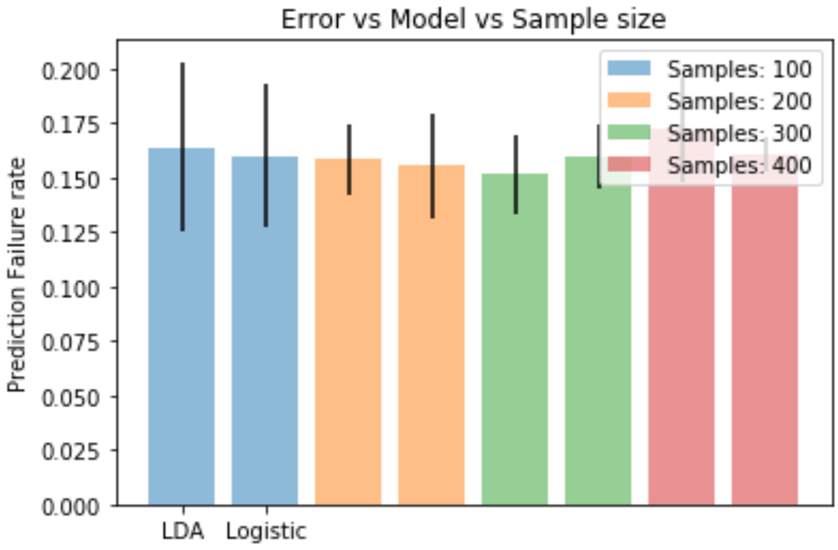
```
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
```

```
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
```

```
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
```

```
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
```

```
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
/Users/ayushsriv/anaconda3/lib/python3.7/site-packages/sklearn/lin
ear_model/logistic.py:433: FutureWarning: Default solver will be c
hanged to 'lbfgs' in 0.22. Specify a solver to silence this warnin
g.
  FutureWarning)
```



## Turn in Instructions

Once you have completed Problems 1 and 2, please submit (for this part of the assignment):

- This .ipynb file.
- A PDF version of this file. To do this:
  1. Go to File -> Download as -> HTML
  2. Open the HTML and Print, and change the **destination** to **PDF**.

In [ ]: