**ntroduction**

Welcome to the assignment on California housing analysis!

Aim:

1. To apply unsupervised dimensionality reduction technique, namely principal component analysis (PCA), to visualize high dimensional data.   
2. To classify samples using linear discriminant analysis (LDA) and multi-layer perceptron  (MLP) classifier, and evaluate the classification performances.

Dataset:

The train and test datasets have been provided below.

**Dataset**

**Download**

Data Description (California Housing Data set)

1. MedInc: median income in block
2. HouseAge: median house age in block
3. AveRooms: average number of rooms
4. AveBedrms: average number of bedrooms
5. Population: block population
6. AveOccup: average house occupancy
7. Latitude: house block latitude
8. Longitude: house block longitude
9. MedianHouseValue: median house value
10. PriceBand: housing classes

This dataset was derived from the 1990 U.S. census, using one row per census block group. A block group is the smallest geographical unit for which the U.S. Census Bureau publishes sample data (a  block group typically has a population of 600 to 3,000 people).

**Comment:**  
Both training and testing datasets have been made available. For PCA, only training data is used as it is just for visualization, while for LDA and MLP both training and testing datasets have to be used. In particular, LDA and MLP classifiers are obtained using training data and their performances compared to testing data.

Tasks:

**A. PCA [40 points]**

1. Use training data housing\_classes\_train.csv. Standardize the data.
2. Choose features (1) to (8). Perform PCA to extract two top features (Principal components).
3. Print number of features, number of samples, number of PC components, and percentage of variance explained by each of these PC components.
4. Print the two eigenvectors used for obtaining the Principal components
5. Obtain a scatter plot of the scores (principal score 2 vs principal score 1)
6. Create a biplot (2 dimensional)
7. State your observations from the biplots

**B. LDA/FDA (Linear Discriminant Analysis or Fisher Discriminant Analysis) [30 points]**

1. Use training data housing\_classes\_train.csv. Standardize the data.
2. Choose features (1) to (8) as the predictor and feature (10) as the target. Build LDA using 2  components.
3. Print number of classes
4. Plot data in the reduced 2 dimensional (LD space)
5. Use test data housing\_classes\_test.csv for prediction.
6. Create a confusion matrix for test data.
7. Print classification report for test data. In particular, print accuracy of the prediction.

**C. MLP Classifier [30 points]**

1. Use training data housing\_classes\_train.csv. Standardize the data.
2. Choose features (1) to (8) as the predictor and feature (10) as the target. Use a multi-layer perceptron classifier. Use early stopping criteria to train the classifier.
3. Use test data housing\_classes\_test.csv for prediction.
4. Create a confusion matrix for test data.
5. Print classification report for test data. In particular, print accuracy of the prediction.
6. Compare the prediction accuracies obtained by the MLP classifier with the LDA classifier and comment on which one is better. Speculate on the reason?

**Report an error**

**Instructions for Submission**

1. File Name: File name should contain assignment number and your email id, e.g., you  are submitting assignment 2 and your email id is xyz@abc.com, please name  submission file as assignment2\_xyz

2. Write your name and email id (full) in the very first block of the code, i.e., the topmost block in your jupyter notebook.

3. File format: After naming the file and executing the code, please convert it to an HTML file.  To convert a jupyter notebook to an HTML file, click on File on notebook, then  Download as and then click on HTML(.html).

4. For jupyter notebook on Colab follow:

* Download your file as .ipynb from Colab.
* Reupload the .ipynb file that you just downloaded to session storage.
* Your file is now uploaded to the current directory. Click on the file icon appearing on the left of the Colab screen and check if your file has been uploaded.
* Suppose the file name is “file\_name.ipynb”, run this command:-  “%% shell jupyter nbconvert --to html file\_name.ipynb” to convert it into .html format.
* Download the .html file.
* For more details refer to the following link which explains the procedure  nicely: https://stackoverflow.com/questions/53460051/convert-ipynb-notebook-to Html-in-google-colab

5. Before converting to HTML make sure all lines of codes are executed and expanded.  This way we will not have to run your code.

6. Before running a model or performing any analysis, please make sure to provide info about the dataset. We want to see the variables and their data types used in the model. For example, if you are going to fit say a linear regression model on a dataset df, simply write df.info() before executing your model.

7. And lastly, please provide all your suggestions and recommendations in one block at the bottom of your code, i.e., it should be your last block of code.

# Submission

This is an individual assignment.  For submissions obtained within one week after the deadline, there will be a 30% penalty. Submissions beyond one week after the deadline will not be accepted.

You have to submit everything in one Jupyter notebook.

Please make sure that you are not changing any of the file names that have been provided for download. The code that you are submitting should run at our end on the same files without any modification of the code.

**Predictive Analysis Assignment**

**Submission details**

**Please Note**: This is an individual assignment. For submissions obtained within one week after the deadline, there will be a 30% penalty. Submissions beyond one week after the deadline will not be accepted.

You need to submit Jupyter notebook.

We strongly recommend you to submit atleast 30 minutes before your deadline

**Module Deadline**

**26 Jun '22**

**12:00 AM (IST)**

Deadline extension can not be requested as due date has crossed.

Upload your file & Save as draft

**Upload**

#### **Note: ONLY Feedback NO Grading post 03 Jul '22, 12:00 AM (IST)**