

## FINAL PROJECT DESCRIPTION

**Team: The Neural Netcases**

**Members: Naomi Hachen, Deepti Panuganti, Anwesha Das**

### FUNCTIONS

#### 1) predict\_glmnet.m

##### **What the function does:**

- 1) Divides data based on city, reassigning cities 1 and 4 to cities 2 and 3, respectively, due to low numbers of sample data.
- 2) Within each city, computes a series of weights across all features based on a downloaded Lasso package, glmnet.
- 3) Make predictions using the weights.

##### **Function Dependencies:**

- 1) It needs the glmnet\_matlab package in the same directory as the .m file.
- 2) The data needs to be in the parent directory (Only the original data provided in project\_kit, no additional data is required).

##### **Running the function:**

The function doesn't require any parameters. It stores the predictions in the variable 'predictions'. It writes the predictions to submit.txt.

#### 2) predict\_PCA\_glmnet.m

##### **What the function does:**

- 1) Divides data based on city, reassigning cities 1 and 4 to cities 2 and 3, respectively, due to low numbers of sample data.
- 2) Runs PCA using to reduce dimensions to 500 principle components using a fast SVD algorithm (using power methods).
- 3) Within each city, computes a series of weights across all features based on a downloaded Lasso package, glmnet.
- 4) Makes predictions using the weights.

##### **Function Dependencies:**

- 1) It needs the glmnet\_matlab package to be in the same directory as the .m file.
- 2) It needs the fsvd.m file to be in the same directory as the .m file.
- 3) The data folder needs to be in the parent directory (Only the original data provided in the project\_kit, no additional data is required).

##### **Running the function:**

The function doesn't require any additional parameters. It stores the predictions in the variable 'predictions'. It writes the predictions to submit.txt.

#### 3) predict\_kmeans\_PCA\_glmnet.m

##### **What the function does:**

- 1) Divides data based on city, reassigning cities 1 and 4 to cities 2 and 3, respectively, due to low numbers of sample data.
- 2) Runs PCA using to reduce dimensions to 500 principle components using a fast SVD algorithm (using power methods).
- 3) Runs kmeans within each city to further subdivide the data.
- 4) Within each kmeans partition, within each city, computes a series of weights across all features and over newly calculated points based on a downloaded Lasso package, glmnet.
- 5) Make predictions using the weights.

##### **Function Dependencies:**

- 1) It needs the glmnet\_matlab package to be in the same directory as the .m file.
- 2) It needs the fsvd.m file to be in the same directory as the .m file.
- 3) The data folder needs to be in the parent directory (Only the original data provided in the project\_kit, no additional data is required).

##### **Running the function:**

The function doesn't require any additional parameters. It stores the predictions in the variable 'prices' and generates a text file 'submit.txt' with the predictions.

#### **4) predict\_PCA\_svm.m**

##### **What the function does:**

- 1) Divides data based on city. NO reassigning is performed.
- 2) Run PCA using to reduce dimensions to 250 principle components using a fast SVD algorithm (using power methods)

##### **Function Dependencies:**

- 1) It needs the libsvm package to be in the same directory as the .m file.
- 2) It needs the fsvd.m file to be in the same directory as the .m file.
- 3) The data folder needs to be in the parent directory (Only the original data provided in the project\_kit, no additional data is required).

##### **Running the Function:**

The function doesn't require any additional parameters. It stores the predictions in the variable 'Y\_hat'. The results will be written to submit.txt.