

Spring 2022: CSEE5590/490 – Special Topics

Python and Deep Learning - ICP-6

Lesson Overview:

In this lesson we will review regression techniques

Regression techniques

- a. Linear Regression
- b. Multiple Regression
- c. Clustering
- d. PCA

Source Code and Dataset:

Provided in the assignment repo.

Regression Assignment:

Q1: Use the same dataset used in the source code (**House_dataset.csv**).

- a. Delete all the outlier data for the GarageArea field (**for the same data set in the use case: House Prices**).

*For this task you need to plot GarageArea field and SalePrice in scatter plot, then check which numbers are anomalies.

- b. Using simple regression select one feature that is positively correlated with 'SalePrice' create a regression model and plot the regression line between the two features.
- c. Find top 5 most correlated features to the target label(SalePrice) and then build a model using those 5 features.
- d. Apply PCA on the same dataset then apply regression model on the PCA result and report your observation after you evaluate the model.
- e. Evaluate all the models using MAE, MSE, RMSE and R2 score.

K-means & PCA Assignment:

Q2: Apply K means clustering to “**credit card dataset: credit_card.csv**”

- a. Remove any null values by the mean.
- b. Use the elbow method to find a good number of clusters with the K-Means algorithm
- c. Calculate the silhouette score for the above clustering.
- d. Try feature scaling and then apply K-Means on the scaled features. Did that improve the Silhouette score?
- e. Apply PCA on the same dataset then apply K-Means algorithm on the PCA result and report your observation if the silhouette score improved or not.

** Follow the IPC rubric guidelines.

Submission Guidelines:

1. Once finished document your code and make sure all parts of the assignments are completed.
2. Push your code to your GitHub repo and update the ReadMe file, add your info, and partner info.
3. Submit the assignment on Canvas.
4. Present your work to TA during class time to prove the execution and complete submission.

After class submission:

1. Once finished document your code and make sure all parts of the assignments are completed.
2. Push your code to your GitHub repo and update the ReadMe file, add your info, and partner info.
3. Submit the assignment on Canvas before the deadline.
4. Record a short video (3~5) minute, proof of execution and complete assignment.
5. Add video link to ReadMe file.

Note: *Cheating, plagiarism, disruptive behavior, and other forms of unacceptable conduct are subject to strong sanctions in accordance with university policy. See detailed description of university policy at the following URL:*
<https://catalog.umkc.edu/special-notices/academic-honesty/>