

CO 544 - Machine Learning and Data Mining

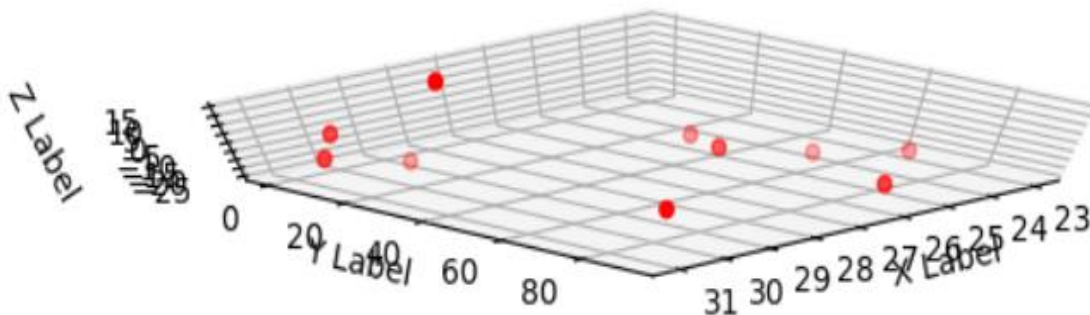
Lab 04

TODO 1: Explain the reason to set, `fig.subplots_adjust(hspace =1.0)` in part (c).

- The reason to set `fig.subplots_adjust(hspace =1.0)` in part (c) is to change the spacing between subplots. The parameter “hspace” is used to set the amount of height reserved for space between subplots to the value 1.0 which is expressed as a fraction of the average axis height.

TODO 2: Visualize the 3D plot in part(e) in a different angle.

- Using the `view_init(60,45)` method to set the elevation and azimuthal angles to visualize the 3D plot in part(e) in a different angle. The elevation of 60 degrees which is 60 degrees above the x-y plane and an azimuth of 45 degrees that is rotated 45 degrees counter-clockwise about the z-axis. The resultant figure is shown below.



Lab Exercise

- Performing Principal Component Analysis (PCA) with 3 components on the wine dataset that is imported from scikit learn standard datasets helps for making data visualization better.
- PCA is affected by scale so first the features need to be scaled in the wine dataset before applying PCA. StandardScaler is used to standardize the wine dataset's features.
- Using PCA, dimension reduction is done and then the plot can be easily understandable. The new components are just the three main dimensions of variations.
- Then, Visualizing the PCA in a 3D plot with well separated class (each of the three classes (0,1,2) are visualized with different colors). Now the classes seem well separated from each other in the graph shown below.

