ESE590 – Lab 01

Asif Iqbal 110333685

The dataset is chosen from UCI ML repository. Dataset “seeds” have 7 features and 210 datapoints in each feature. Each of the features represents a property of wheat seed. There are 3 different kinds of wheat seeds: Kama, Rosa and Canadian. Clustering and classification algorithms can be applied to this dataset.

Scatter Plot:

Plots the datapoints in a 2D grid. We plotted a scatter plot matrix. It plots each feature against each other. The diagonal of the matrix is the kernel density estimation plot of the 3 types of wheat seeds.

Background pattern

Description automatically generated

Figure 1: Scatter Matrix of seeds dataset

KDE Plot:

Estimates and plots the probability density function of the passed in dataset. This plot does not assume an underlying distribution for the dataset.

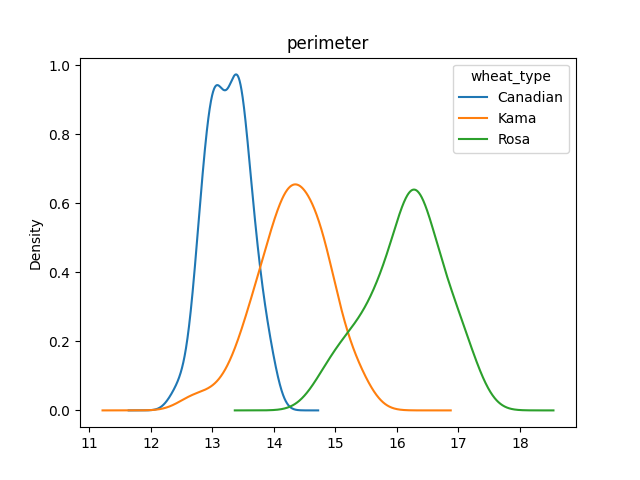


Figure 2: KDE plot of feature “parameter”

Strip Plot:

Deals with collision of datapoints. If multiple data points are mapped to the same coordinate, the latter ones are not visible. Strip plot gets around this issue. If a point already exists in a mapped coordinate for a new point, it randomly moves the new point either little above or little below the existing point. This method is called “jitter”.

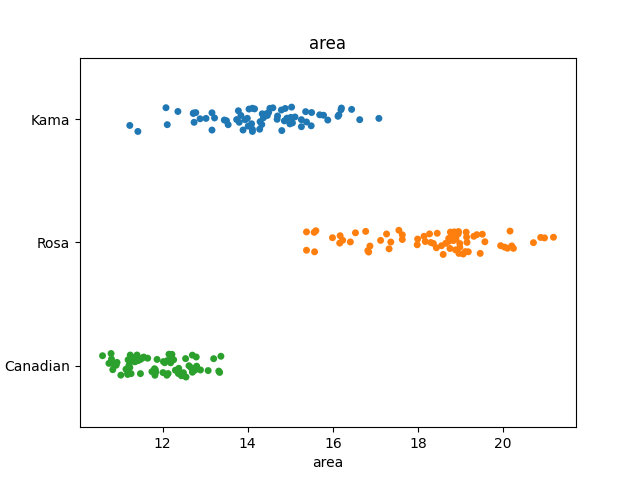


Figure 3: Strip plot of feature “area”

Swarm Plot:

A picture containing icon

Description automatically generated Sorts the data values and maps it on a grid. If the values are all “1”, on the 1D grid, it will map the 1’s on top of another.

Chart, scatter chart

Description automatically generated

Figure 4: Swarm Plot of feature “compactness”

Box Plot:

Sub divides the data where 50% is on the right and 50% on the left. Subdivide the left and right again in the same fashion. There are 3 lines now, which represents 25th percentile, median and 75th percentile of the dataset. The whiskers represent the min and max value in the dataset. And the diamond shaped dot represents the outliers.

Chart, box and whisker chart

Description automatically generated

Figure 5: Box plot of feature “kernel\_length”

Bar Plot:

Useful while working with categorical feature. The “wheat\_type” feature is a categorial type.

Chart, bar chart

Description automatically generated

Figure 6: Bar plot of feature “kernel\_width”

Takes the mean of the kernel width associated with each wheat type and represents as a bar on the plot.

More plots can be found by running the source code. There is a function for each plot. Calling the functions will generate specific plots.

Dataset URL: <https://archive.ics.uci.edu/ml/datasets/seeds>