Lab Assignment– Chapter 3: Model Training and Evaluation

Q1. Data Holdout

Load the dataset auto-mpg.csv and perform the following tasks:

1. Print the shape of the dataset.
2. Split the dataset into training (80%) and testing (20%) subsets.
3. Use a random state for reproducibility.
4. Print the training and test sets and their length.

Q2. Implementing 5-Fold Cross-Validation

Using the dataset btissue.csv, do the following:

1. Create a 5-fold cross-validation.
2. For each fold (starting with 0), print:
   - fold number
   - indices of the training set
   - indices of the test set
3. Print the training set and test set and their length for the 3rd iteration.

Q3. Bootstrap Sampling

With the dataset btissue.csv, perform the following:

1. Extract the predictors.
2. Generate a bootstrap sample of size 100.
3. Display the first 10 rows of the resampled data.

Q4. Random Forest Classification and Evaluation

Using the dataset btissue.csv, carry out the steps below:

1. Separate predictors and target.
2. Perform a holdout split (80% training, 20% testing).
3. Train a Random Forest Classifier.
4. Evaluate the model using:
   - accuracy_score
   - confusion_matrix

Q5. Regression and Evaluation

Load the dataset auto_mpg.csv and perform the following tasks:

1. Remove all rows where value of any column is ‘NaN’.
2. Separate predictors and target.

3. Perform a holdout split (80% training, 20% testing).
4. Train a Linear Regression Model.
5. Evaluate the model using:
   - mean_squared_error
   - r2_score

Q6. Clustering and Quality Measures

Using the dataset Dataset_spine.csv:

1. Exclude the class (last) column to form predictor variables.
2. Apply K-Means clustering with n_clusters=3 and random_state=123.
3. Compute and print:
   - v-measure score
   - Silhouette score