## 1. Fruits-360 Classification

## 1) Dataset Information:

- Dataset Name: Fruits-360 Original Size

- Feature vector size: 50 × 50 × 3 = 7500 features per image

- Total Samples Used: 2300

- Size of each image: 50 × 50 pixels, RGB

**- Number of classes used:** 5 classes.

- Labels: Extracted automatically from dataset folders (e.g., Apple 10, Apple 11, Apple 12, etc.)

- Dataset split:

Training set: 80% (1840 images)

Testing set: 20% (460 images)

Validation set: Not used

## 2) Implementation Details

## **A) Feature Extraction:**

- Each image is resized to 50×50 pixels and converted to RGB format.

- Images are flattened into a one‑dimensional vector of raw RGB pixel intensities.

- Pixel values are normalized to the range [0,1].

- Extracted features: Raw RGB pixel values

- Number of extracted features: 7500

- Feature dimension: (7500,)

1. **Logistic Regression Hyperparameters :**

Model type: Multiclass Logistic Regression

Regularization: L2 (default)

Maximum number of iterations: 900

Optimizer (for manual loss curve): Gradient Descent

Learning rate: 0.1

Number of epochs: 150

1. **K‑Means Hyperparameters:**

Number of clusters (k): Equal to number of classes

Initialization: k‑means++ (default)

Random state: 42

## 3) Results

Included Metrics:

#### Accuracy

Accuracy represents the percentage of correctly classified test samples.

Logistic Regression Accuracy is: 100%

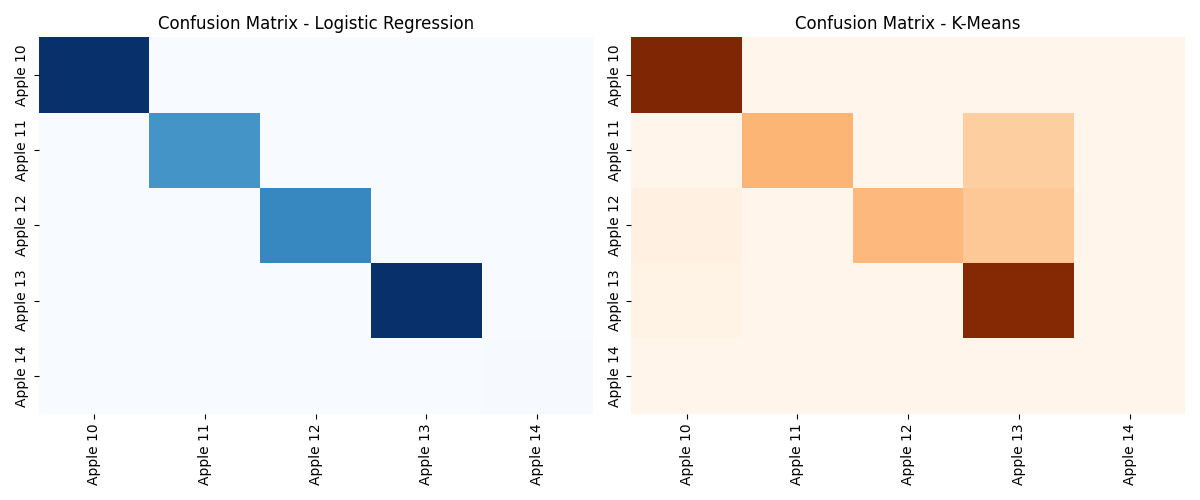
K means Accuracy is: 81%





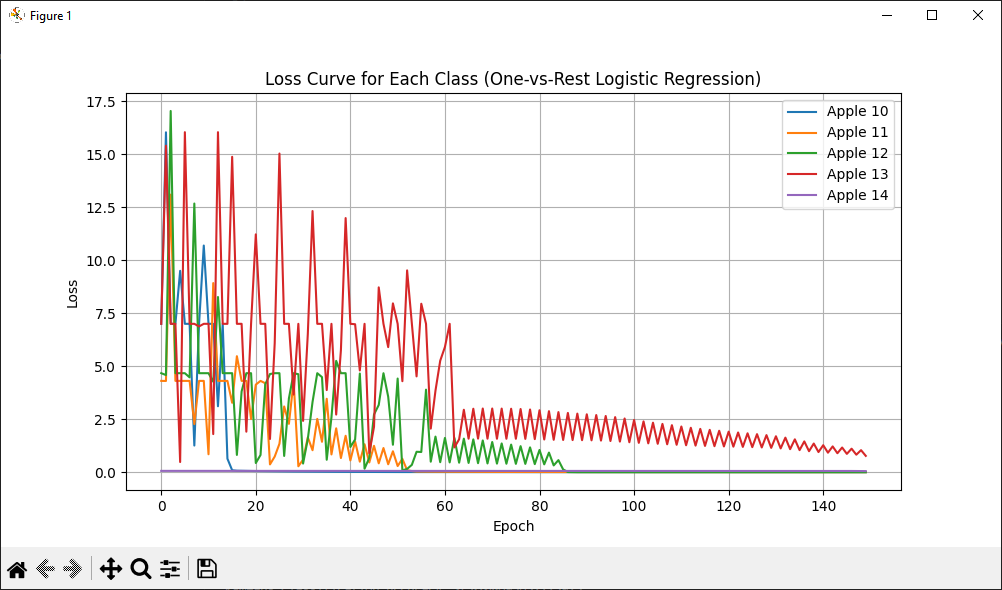
#### Confusion Matrix

The confusion matrix shows the number of correct and incorrect predictions for each class. Diagonal values indicate correct classifications, while off‑diagonal values represent misclassifications between classes.



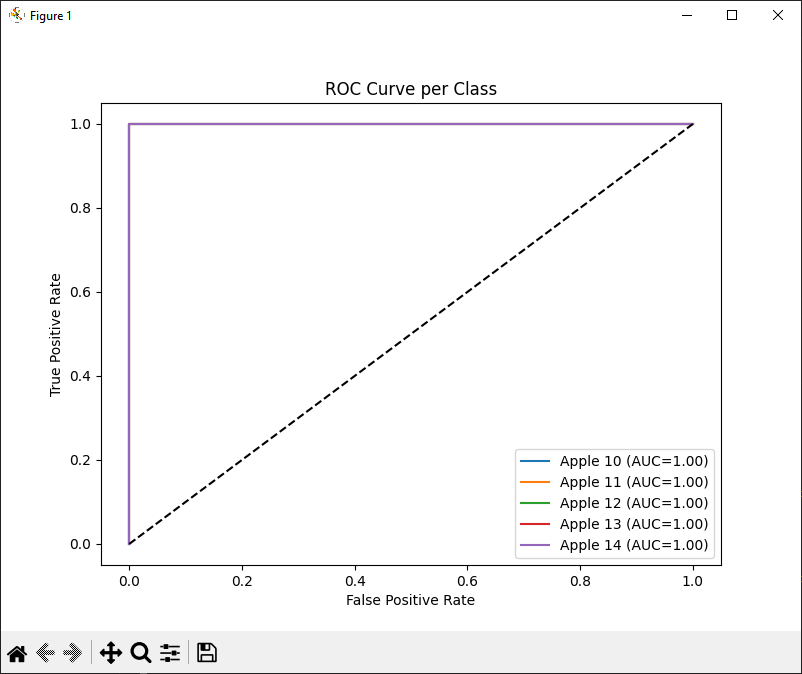
#### Loss Curve

The loss curve illustrates how the logistic regression model minimizes the cross‑entropy loss over training epochs using gradient descent. A decreasing loss curve indicates effective learning and convergence of the model.



#### ROC Curve and AUC

Receiver Operating Characteristic (ROC) curves were generated using a One‑vs‑Rest strategy for multiclass classification. The Area Under the Curve (AUC) measures class separability, higher AUC values indicate better classification performance.



## 2. Insurance Prediction

## 1) Dataset Information

Dataset: insurance.csv

Features: age, bmi, children, smoker, region...

Target: charges

No missing values

One-hot encoding applied

Train/Test Split: 80% / 20%

## 2) Implementation Details

Scaling: StandardScaler

Linear Regression: Ordinary Least Squares

KNN: n\_neighbors = 5

## 3) Results

Metrics:

- MSE

- RMSE

- R² Score



**Plot: Actual vs Predicted (LR and KNN)**

