Assignment: Image Classification Using Logistic Regression

Scenario:

You are working as a Machine Learning Engineer for a wildlife research organization. The team wants to develop a simple binary image classification model that can distinguish between **rabbits and squirrels** captured on wildlife cameras in the forest.

Due to computational constraints, the team wants to begin with a **logistic regression** model before moving to complex architectures like CNNs.

Your goal is to build a complete logistic regression pipeline from scratch using image data — including **data loading, preprocessing, model training using gradient descent, evaluation using classification metrics**, and finally a **cost vs iteration plot**.

Dataset:

We'll use the **"Rabbits vs Squirrels"** dataset hosted on Kaggle.

Download Dataset from Kaggle

After downloading:

- Extract the dataset in a folder (e.g., D:/ML_Assignment/rabbits_vs_squirrels/)
- The folder should have:
    - train/rabbit/, train/squirrel/
    - test/rabbit/, test/squirrel/

Assignment Tasks

Task 1: Data Preprocessing

1. Load the images from both folders.
2. Convert all images to grayscale and resize them to a smaller size (e.g., 64x64).
3. Normalize pixel values and flatten images into vectors.
4. Assign labels: rabbit = 0, squirrel = 1.

Task 2: Build Logistic Regression from Scratch

1. Initialize weights and bias.

2. Implement the sigmoid function.
3. Compute the binary cross-entropy loss.
4. Use gradient descent to optimize the parameters.

Task 3: Training and Evaluation

1. Train the model for multiple iterations.
2. Plot **cost vs iterations**.
3. Evaluate the model on the test set using:
    - Accuracy
    - Confusion matrix
    - Precision, Recall, and F1-score

Task 4: Visualization

1. Show 10 random test images with predicted and true labels.
2. Plot a heatmap of the confusion matrix using Seaborn.

Deliverables:

1. Python notebook or .py script with code and comments.
2. Screenshots of output:
    - Cost vs iteration graph
    - Confusion matrix
    - Sample predictions
3. A short PDF report answering the **assignment questions** below.

Assignment Questions

1. Explain the importance of normalization in image data before training a model.
2. Why is logistic regression suitable for binary classification tasks?
3. What are the limitations of logistic regression for image data?

4. How does gradient descent update the model parameters?
5. How can we improve accuracy further after logistic regression?