Questions

Model performance metrics

Linear Discriminative analysis:

Train Accuracy of LDA on RGB Version : 0.51466

Test Accuracy of LDA on RGB Version : 0.3713

Train Accuracy of LDA on Greyscale Version : 0.35164

Test Accuracy of LDA on Greyscale Version : 0.2785

Quadratic Discriminant Analysis:

Train Accuracy of QDA on RGB Version : 0.99244

Test Accuracy of QDA on RGB Version : 0.3623

Train Accuracy of QDA on Greyscale Version : 0.88622

Test Accuracy of QDA on Greyscale Version : 0.4473

Gaussian Naive Bayes:

Train Accuracy of GNB on RGB Version : 0.28472

Test Accuracy of GNB on RGB Version : 0.2976

Train Accuracy of GNB on Greyscale Version : 0.2583

Test Accuracy of GNB on Greyscale Version : 0.268

1. State the maximum likelihood estimates for the parameters of Gaussian Naive Bayes. Do these estimates change if we use grayscale images instead of RGB images? Explain why or why not.

The maximum likelihood estimates for the parameters in Gaussian Naive Bayes Model are mean and variance of the pixels in each class

- Mean $(\mu) = (\Sigma x) / N$, where Σx is the sum of values and N is the number of images. *

- Variance $(\sigma^2) = (\Sigma(x - \mu)^2) / N$, where $(x - \mu)$ is the deviation of each value from mean and N is the number of images. *

In RGB each pixel has three values red, green and blue while the greyscale pixel has only one value but the parameter estimates are measured independently for each class. So estimates do not change if we use grayscale images instead of RGB images

2. The accuracy of QDA using RGB images was lower than that of grayscale images. What assumptions does QDA make that might cause

this difference in performance?

Below are the assumptions that might cause difference in QDA model's performance between RGB images and greyscale images resulting in lower accuracy on RGB images

- The pixel values follow multivariate normal distribution for each class*
    - RGB images have three features (Red, Green and Blue) in each pixel while the grayscale images have one feature defining the intensity of each pixel.so the QDA model will face complexity incase of RGB because of higher dimentional feature space (R,G,B features)
- Equal Covariance Matrices (assumes each class has its own covariance matrix)*
    - In RGB images, matrices become complex because of higher dimentional feature space (R,G,B features) and might result in overfitting
- QDA is sensitive to sample size*
    - As RGB images have higher dimensions (R,G,B features) the QDA model needs more training samples for beeter performance when compared to grayscale images. this might result in lower accuracy of QDA model on RGB images than that of grayscale for same sample size

3. Both LDA and Gaussian Naive Bayes saw reduced test accuracy on grayscale images compared to RGB images. Why might this be the case (is it the data, the model, or something else)?

- RGB images contain more information in terms of Red, Green, Blue features in each pixel value and Grayscale images contain only one feature defining the intensity of each pixel.
- Converting RGB images to grayscale images results in loss of information and drop in accuracy
- LDA and GNB models might have faced complexity with loss of information in grayscale images, while QDA performed better on grayscale by modeling covariance matrices. As LDA and QDA ignores the assumption of equal covariance matrices across classes

4. How many parameters are estimated for each model and each image type (RGB and grayscale)?

Linear Discriminant Analysis:

- **RGB Images:*** 39 parameters
    - Mean: 30 parameters (3 channels per class, 10 classes)
    - Covariance Matrix Shared: 9 parameters (3x3 matrix)
- **Grayscale Images:*** 11 parameters
    - Mean: 10 parameters (1 per class)
    - Covariance Matrix Shared: 1 parameter

Quadratic Discriminant Analysis:

- **RGB Images:*** 120 parameters
    - Means: 30 parameters (3 channels per class, 10 classes)
    - Covariance Matrices: 90 parameters (3x3 matrix per class, 10 classes)
- **Grayscale Images:** Total: 20 parameters
    - Means: 10 parameters (1 per class)
    - Covariance Matrices: 10 parameters (1x1 matrix per class, 10 classes)

Gaussian Naive Bayes:

- **RGB Images:*** 70 parameters
  - Priors: 10 parameters
  - Means: 30 parameters (3 channels per class, 10 classes)
  - Variances: 30 parameters (3 channels per class, 10 classes)
- **Grayscale Images:*** 30 parameters
  - Priors: 10 parameters
  - Means: 10 parameters (1 per class)
  - Variances: 10 parameters (1 per class)

In [ ]: