CSE6363: Machine Learning
Assignment-2 Question and Answers

Question 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.

Answer: The maximum likelihood estimates for Gaussian Naive Bayes are means and variances calculated for each class in each feature

1. **Mean (μ):** The mean is calculated for each class c by the average values of each feature d belonging to that class group

$$\mu_{cd} = \frac{1}{N_c} \sum_{k=1}^{N_c} x_{k,d}$$

where N_c is the number of samples in class c and $x_{k,d}$ is the value of feature d in the k-th sample of class c.

2. **Variance (σ^2):** The variance is calculated for each class is the average value of the sum of the square of deviation of each feature d values from its class mean value:

$$\sigma^2_{cd} = \frac{1}{N_c - 1} \sum_{k=1}^{N_c} (x_{k,d} - \mu_{cd})^2$$

where N_c is the number of samples in class c, μ_{cd} is the mean of feature d in class c, and $x_{k,d}$ is the value of feature d in the k-th sample of class c.

Above estimates **do not change** if we use greyscale images instead of RGB. This is because means and variances are calculated individually for each pixel considering the features independently. The value of the greyscale in a single pixel will be the average value of the RGB when converting from RGB to greyscale.

Question 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?

Answer: QDA (Quadratic Discriminant Analysis) is estimated based on the means and covariance matrix values of each class and assumes each class's conditional density is Gaussian. In the case of grayscale images, each pixel has one channel with a value between 0 and 255. The distribution of these grayscale channel values will follow Gaussian. But in the case of RGB, each pixel will have 3 channels representing intensity levels of red, green, and blue, and together, these 3 channels might not follow the Gaussian distribution. And also, since the number of dimensions increased for RGB compared to grayscale, this might lead to a dimensionality issue. Because of these factors, it might lead to a drop in the accuracy of QDA using RGB.

Question 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)?

Answer: Accuracy on grayscale images compared to RGB images is seen less in LDA, and Gaussian Naive Bayes might be because of the below reasons.

- LDA assumes **shared covariances** and Naive Bayes assumes **feature independence**. Because of these assumptions, we may not be able to calculate optimal performance.
- Greyscale images only contain the data to distinguish between the classes, but they will discard the color data, which might be beneficial for identifying the separation between the classes. Whereas RGB has more data compared to grayscale, which might help in better classifying the images.

Question 4

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

Answer:

1. **LDA** : It estimates the mean for N number of features for each of the C classes. It results in $N * C$ parameters for greyscale and $3 * N * C$ parameters in case of RGB. LDA assumes equal covariance matrices for all classes, resulting in a single shared covariance matrix for both RGB and greyscale. It also produces one weight and bias parameters using the determined class mean and shared variance.
2. **QDA**: similar to LDA, it also estimates the mean for N number of features for each of the C classes. It results in $N * C$ parameters for greyscale and $3 * N * C$ parameters for RGB. QDA assumes separate covariance matrices for each class with the dimension $N * N$ for greyscale and $3 * N * N$ parameters for RGB.

3. **Naive Bayes**: Naive Bayes assumes feature independence. Thus, the mean and variance are computed for each pixel in each class. Parameters for grayscale $2 * N * C$ and RGB $3 * 2 * N * C$ are estimated for both mean and variance, respectively.