PAVAN BABU. B
1002207463
ML-ASSIGNMENT-2

ASSIGNMENT-2 → QUESTIONS

1) The maximum likelihood estimates for Gaussian Naive Bayes are:

(i) Mean (μ): Estimated as the sample mean for each features in class.

(ii) Variance (σ^2): Estimated as the sample variance for each feature in each class.

The estimates fundamentally doesn't change but the no. of features changes when using grayscale instead of RGB.

→ RGB images have 3 channels, so there are 3 sets of parameters (μ, σ^2) per pixel

→ Grayscale images have 1 channel, so there is 1 set of parameters (μ, σ^2) per pixel.

The estimation process remains the same, but it's applie to fewer features in the grayscale.

(2) QDA performance: RGB vs Grayscale.

QDA's lower accuracy with RGB images compared to grayscale might be due to its assumptions:

=> Multivariate Gaussian Distribution

QDA assumes features follow a multivariate Gaussian distribution for each class.

=> Different covariance Matters:

QDA allows each class to have its own Co-variance matrix.

With RGB images, QDA needs to estimate more parameters, potentially leading to overfitting, especially with limited data.

Grayscale images, having fewer features may provide a better balance between model complexity and available data.

(3) LDA and Gaussian Naive Bayes: Reduced accuracy on Grayscale

The reduced accuracy for LDA and Gaussian Naive Bayes on grayscale images could be due to:

(i) Loss of information: Grayscale conversion loses color information that might be crucial for classification

(ii) Feature Importance: Colour features in RGB might be more discriminative for the specific classification task.

(iii) Model assumptions: Both models assume Gaussian distributions, which might fit RGB data better in this case.

-> The outcome is task-specific. In some cases, grayscale images perform better, especially when color is not a critical factor for classification.

(4). Parameter Estimation for each model and Image type.

Comparing the no. of parameters estimated for each model and image type:

Model	RGB Parameters	Grayscale Parameters
Gaussian Naive Bayes	$6n+c$	$2n+c$
LDA	$3n+3n(3n+1)/2+c$	$n+n(n+1)/2+c$
QDA	$c(3n+3n(3n+1)/2)$	$c(n+n(n+1)/2)$

Where: $n \rightarrow$ no. of pixels; $c \rightarrow$ no. of classes.

GNB estimates fewer parameters due to its independence assumption. LDA estimates more parameters as it computes as shared covariance matrix. QDA estimates the most parameters as it computes a seperate covariance matrix for each class.