1 Questions

1.1 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.

1.1.1 MLE for parameters

In Gaussian Naive Bayes, the parameters we need to estimate are class priors(π_k) ,mean(μ_k) and variance(σ_k^2) for each class k and each feature n

- **Class Priors(π_k)**

$$\pi_k = \frac{N_k}{N} \tag{1}$$

where N_k is the number of samples in class k, and N is the total number of samples.

- **class mean(μ_k)**

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^{N_k} x_i \tag{2}$$

where N_k is the number of samples in class k, and x_i is the i^{th} sample in class k.

- **class variance(σ_k^2)**

$$\sigma_k^2 = \frac{1}{N_k} \sum_{i=1}^{N} (X_i - \mu_k)^2 \tag{3}$$

where N_k is the number of samples in class k, and x_i is the i^{th} sample in class k, and μ_k is the mean of the samples in class k.

1.1.2

these estimates will change if we use grayscale images instead of RGB images.

Reason

The calculation of mean and variance is based on the feature of the data which is affected by the amount of data and feature dimension.There are three features per pixel (red, blue and green) in RGB image whereas grayscale image has only one feature per pixel, hence using grayscale map reduces the feature vector dimension. Since the amount of data is constant and reduces the feature dimension, it reduces the mean and variance.

1.2 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?

- **Assumption of Normal Distribution of Features**:

  QDA assumes that the features of each class follow a multivariate normal distribution. In RGB images, three color features in each pixel may not normally distributed, which may cause the feature distribution to deviate from the normal distribution and affect the performance of QDA.

- **Different covariance matrix assumptions for each class**:

  QDA assumes that each class has a different covariance matrix. In RGB images, there may be a strong correlation between the three color features in each pixel, resulting in the covariance matrix becoming complex and difficult to estimate, which directly affects the performance of QDA.

1.3 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)?

- **Data**: RGB images have three features (red, green and blue) per pixel, while grayscale images combine these three features into one, losing information in the data, especially color information.

- **Model**: Both models, LDA and Gaussian Park Bayes, rely on the assumption of feature distribution, and in grayscale images, the reduction of features affects the performance of the model.

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

parameters estimated for each model and each image type are shown in the figure below.

mode	**data**	**parameter**	**counts**	**total**
LDA	RGB	Mean Covariance Class Priors	3072 3072^2 10	9,440,266
	grayscale	Mean Covariance Class Priors	1024 1024^2 10	1,049,610
GDA	RGB	Mean Covariance Class Priors	3072 $3072^2 \times 10$ 10	94,374,922
	grayscale	Mean Covariance Class Priors	1024 $1024^2 \times 10$ 10	10,486,794
GNB	RGB	Mean Variance Class Priors	3072 3072^2 10	9,440,266
	grayscale	Mean Variance Class Priors	1024 1024^2 10	1,049,610

Tabelle 1: Number of parameters estimated for model