PROJECT PART -1:
DENSITY ESTIMATION AND CLASSIFICATION USING FASHION-MNIST

OBJECTIVE:

The project focusses on developing two machine learning models - Naive Bayes Classifier and Logistic Regression. The steps involved are as follows:

1. Calculate two features of each image - average and standard deviation.
2. Estimate parameters for 2D Gaussian distribution for each class.
3. Implement Naive Bayes to perform classification.
4. Train Logistic regression using gradient ascent to perform classification.

DATASET:

The dataset has training set and testing set. The number of samples in training set are 12000 and number of samples in testing set are 2000.
Number of samples in the training set: “Tshirt”:6000; “Trouser”:6000.
Number of samples in the testing set: “Tshirt”:1000; “Trouser”:1000.

FEATURES:

We are extracting the following two features for each image in training and testing dataset:

- The average of all pixel values in the image.
- The standard deviation of all pixel values in the image.

The training dataset is being converted from 12000*784 is getting converted to 12000*2

```
fet1=np.mean(trainX,axis=1)
fet2=np.std(trainX,axis=1)
trainX=pd.DataFrame({'trainF1': fet1, 'trainF2': fet2})
fett1=np.mean(testX,axis=1)
fett2=np.std(testX,axis=1)
testX=pd.DataFrame({'testF1': fett1,'testF2':fett2})
```

In the above code,
trainX represents training dataset.
trainF1 represents the feature 1 of the training dataset.
trainF2 represents the feature 2 of the training dataset.
testF1 represents the feature 1 of the testing dataset.
testF2 represents the feature 2 of the testing dataset.

ESTIMATION OF PARAMETERS:

Parameters are estimated using the below formulae (mean and standard deviation)
We calculate standard deviation(σ) from variance.

$$\mu = \sum_{i=1}^{n} x_i / n$$

$$\sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2 \, / \, n - 1$$

The values

CLASS	Feature1(mean)	Feature1(std)	Feature2(mean)	Feature2(std)
CLASS 0	0.325608	0.113375	0.320036	0.087983
CLASS 1	0.222905	0.056951	0.333942	0.057032

NAÏVE BAYES CLASSIFICATION:

We are using 1D gaussian distribution equation to compute the probability. The equation is

$$\sum_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} \; e^{\frac{-(x_i - \mu)^2}{2\sigma^2}}$$

Where μ is average; σ is standard deviation; x_i is sample.

According to bayes theorem the equation for computing probability of the data is

$$P(Y|X) = \frac{P(X|Y) * P(Y)}{P(X)}$$

$$P(Y|X) = \frac{P(X_1|Y) * P(X_2|Y) * P(Y)}{P(X)}$$

Where P(Y) is prior, which can be calculated using the formula.

P(Y)=ratio of samples belonging to a class to total number of samples.

The predicted probability can be calculated by using the formulae

$$argmax_Y P(Y|X) = argmax_Y P(X_1|Y) * P(X_2|Y) * P(Y)$$

$$argmax_Y P(Y|X) = argmax_Y \log\big(P(X_1|Y)\big) + \log\big(P(X_2|Y)\big) + \log\,(P(Y))$$

We compute the probabilities of sample belonging to class 0 and class1 and if P(Y=0/X) is greater than P(Y=1/X) then sample belong to class 0 otherwise it belongs to class1.

The accuracies computed are shown below

```
Class Tshirt accuracy : 78.40 %
Class Trouser accuracy: 87.90 %
Overall accuracy      : 83.15 %
```

LOGISTIC REGRESSION CLASSIFICATION:

To optimize logistic regression, we use likelihood estimation and the goal here is to maximize the likelihood we can achieve this through the Gradient ascent algorithm.

We use sigmoid function to calculate the output probabilities.

$$z = \sigma(w \cdot X + w\_0)$$

$$\sigma(t) = \frac{1}{1 + e^{-t}}$$

Where w is weights; X is sample and w_0 is bias parameter.

The equation for log likelihood is as follows:

$$L(w, w\_0) = \sum_{i=1}^{n} y^{(i)} \cdot \log\big(z^{(i)}\big) + (1 - y^{(i)}) \cdot \log\,(1 - z^{(i)})$$

The gradient of log likelihood is derivative of log likelihood function. The equation is

$$\frac{\partial L(w)}{\partial w} = (y - z) \cdot X$$

$$\frac{\partial L(w\_0)}{\partial w\_0} = \sum (y - z)$$

The weights are updated by using the above derivatives.

$$w = w + \eta * \frac{\partial L(w)}{\partial w}$$

$$w\_0 = w\_0 + \eta * \frac{\partial L(w\_0)}{\partial w\_0}$$

The weights obtained from gradient ascent are -177.69004367 and 189.13702829

The bias value is -14.804176432569857

Using weights obtained from the training data, we compute the predicted values using learning rate as 0.01and number of iterations as 500. These predicted values are classified using a threshold. The threshold value used is 0.5.

If the predicted values are greater than 0.5, it is classified as Trouser class otherwise T-shirt class.

In this way the dataset has been classified and the accuracies are obtained as below.

```
Class Tshirt accuracy : 92.50 %
Class Trouser accuracy: 91.70 %
Overall accuracy      : 92.10 %
```