

Nama : Hafidh Fikri Rasyid

NIM : 1301142190

Kelas : IF-38-10

### Assignment 3

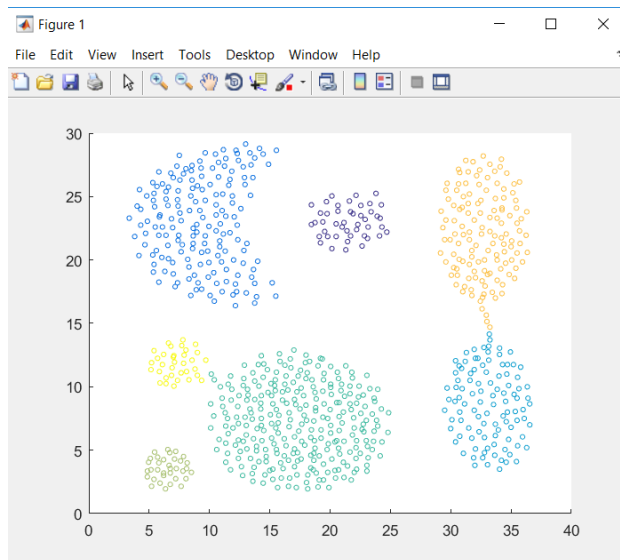
#### Machine Learning

#### Soal

1. Create a function to calculate performance of classifier that implements F1-micro average, F1-macro average, and simple accuracy (i.e. number of correct prediction divided number of data). (Hints: give an input parameter of the function that inform the function when it should use either F1-micro average, F1-macro average, or simple accuracy). **[5 points]**
2. **(45 points)** In this problem we implement Naive Bayes for classification.
  - (a) Load the selected data set. Visualize all data points using scatter plot. From your scatter plot, one could easily distinguish each class. (hint: Use attribute 1 as x -axis, attribute 2 as y -axis. Use different color and/or different symbol for each class label).**[5 points]**
  - (b) Apply Naive Bayes classifier on the selected data set. Your codes have to clearly contain
    - i. Function for learning that implements the calculation of prior and likelihood probabilities. **[10 points]**
    - ii. Function for predicting/classifying data that implements the calculation of posterior probabilities. It had better for you to implement the logarithm (log) function.**[10 points]**
    - iii. Classify each data point using the trained Naive Bayes. Plot the results using scatter plot. From your scatter plot, one could easily distinguish each class. (hint: use different color and/or different symbol).**[5 points]**
    - iv. By visually comparing figures created from point 2(a) and 2(b)iii, what do you think of the classification results using Naive Bayes. **[5 points]**
  - (c) Use the function of exercise 1 to evaluate performance of Naive Bayes classifier. What is the F1-micro average of Naive Bayes classifier on the data set. **[5 points]**
  - (d) Plot the decision boundary resulted from Naive Bayes classifier on the figure that has been created by point 2(a). (Hints: generate data points using range of minimum and maximum value of each attribute, then classify each generated data points using trained Naive Bayes classifier. Use attribute 1 and attribute 2 as both x -axis and y -axis of decision boundary location, while the predicted class label for coloring).**[5 points]**

#### Jawaban

1. Sebuah fungsi yang saya buat adalah untuk menghitung akurasi, F1-Macro Average Score dan F1-Micro Average score dari inputan confusion matrix dari semua kelas yang berisi True Positif, True Negatif, False Positif dan False Negatif. Untuk fungsi diatas dapat dilihat pada file F1AndAccFunction.m pada aplikasi yang telah dikirimkan.
2.
  - a. Data dataset yang ditampilkan adalah seperti dibawah ini



- b. Untuk main program dari aplikasi ini terdapat pada file NaiveB.m pada aplikasi yang telah dikirimkan.
- i. Untuk fungsi yang berfungsi untuk menghitung prior dan likelihood dapat dilihat pada file hitungPrior.m untuk menghitung prior dan hitungLikelihood.m untuk menghitung likelihood. berikut adalah rumus mencari prior.

$$\text{Prior} = \frac{x}{n}$$

Dari rumus diatas dapat dijelaskan bahwa x adalah banyak data pada label tertentu dan n adalah banyak data secara keseluruhan. Untuk implementasi kode dari rumus diatas adalah sebagai berikut :

```
function prior = HitungPrior(data,n,lbl)
    prior = sum(data(:,3)==lbl)/n;
end
```

Dibawah ini adalah rumus mencari likelihood

$$\text{Likelihood} = \frac{1}{\sqrt{2\pi}\sigma_{ji}} \exp\left(-\frac{(X_j - \mu_{ji})^2}{2\sigma_{ji}^2}\right)$$

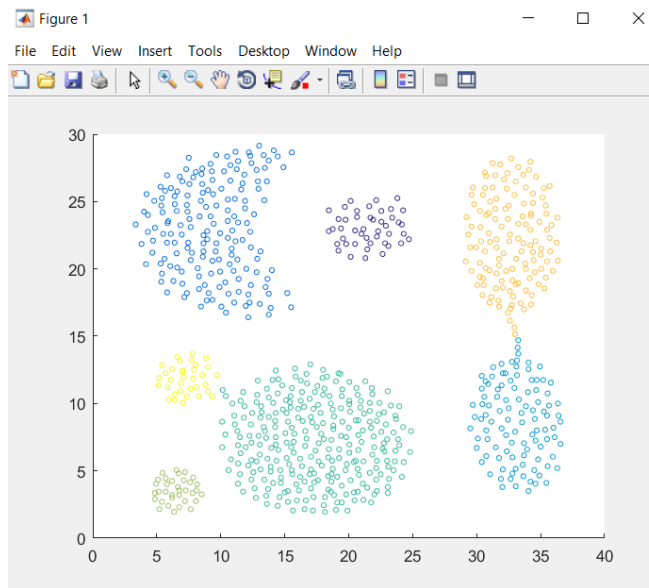
Dari rumus diatas dapat dijelaskan bahwa  $X_j$  adalah data yang akan dihitung likelihoodnya berdasarkan suatu kelas lalu  $\sigma_{ji}$  adalah standar deviasi dari sekumpulan data berdasarkan kelas tertentu dan  $\mu_{ji}$  adalah nilai rata-rata dari sekumpulan data berdasarkan kelas tertentu. Untuk implementasi kode dari rumus diatas adalah sebagai berikut :

```
function likelihood = HitungLikelihood(meanData,stdData,x)
    likelihood = (1/(stdData*sqrt(2*pi)))*exp(-(x-meanData)^2/(2*(stdData)^2));
end
```

- ii. Untuk fungsi yang berfungsi untuk menghitung posterior terdapat pada file hitungProst.m pada aplikasi yang telah dikirimkan. Untuk menghitung posterior yaitu dengan mengalikan masing-masing likelihood dari setiap data dan dikalikan juga dengan likelihood.

```
function prosterior = HitungProst(l1,l2,prior)
    prosterior = l1*l2*prior;
end
```

- iii. Gambar scatter plot hasil dari klasifikasi menggunakan metode naïve naves dapat dilihat dibawah ini :



- iv. Dari dua scatter plot yang telah ditampilkan diatas bahwa hampir tidak ada perbedaan distribusi data antara data original dengan hasil klasifikasi naïve bayes. Ini terjadi Karena data yang ada cenderung berkumpul pada satu kelompok. Hal ini menyebabkan data mudah untuk diklasifikasikan oleh naïve bayes.
- c. Hasil F1-Micro Average dari klasifikasi naïve bayes diatas yang dihitung menggunakan fungsi pada nomer 1 adalah sebesar 99.8731%.

Command Window

```
F1 Micro Average dari dataset diatas adalah 99.8731%
```

```
fx >>
```

- d.