Tugas 1 EB4011 Pengenalan Pola

Dataset yang digunakan dalam tugas ini adalah dataset bunga “*Irish*”. Dalam dataset tersebut terdapat 3 kelas output klasifikasi, namun dalam tugas ini hanya digunakna 2 kelas output klasifikasi. Kedua output tersebut diambil dari 2 output pertama dalam dataset, yaitu bunga *Irish Setosa* dan bunga *Irish Versicolor*. Masing-masing dari kedua dataset bunga *Irish* tersebut berjumlah 50 data sehingga total ada 100 data. Dalam dataset tersebut digunakan empat parameter untuk menentukan kelas outputnya, antara lain panjang petal (*petal length*), lebar petal (*petal width*), panjang sepal (*sepal length*), dan lebar sepal (*sepal width*). Keempat parameter tersebut dieksplorasi dengan melakukan plotting menjadi 2, yaitu berdasarkan panjang & lebar dari sepal dan petal.

from sklearn import datasets

import numpy as np

import matplotlib.pyplot as plt

#load dataset

iris = datasets.load\_iris()

x = iris.data[:100]

y = iris.target[:100]

#diviide sepal and petal matrix

x\_sepal\_length = x[:, [0]]; x\_sepal\_width = x[:, [1]]

x\_petal\_length = x[:, [2]]; x\_petal\_width = x[:, [3]]

#plotting sepal

plt.plot(x\_sepal\_length[:50], x\_sepal\_width[:50], 'o', color='blue', label='setosa')

plt.plot(x\_sepal\_length[50:], x\_sepal\_width[50:], 'x', color='red', label='versicolor')

plt.title('plotting sepal')

plt.xlabel('sepal\_length')

plt.ylabel('sepal\_width')

plt.legend(); plt.show()

#plotting petal

plt.plot(x\_petal\_length[:50], x\_petal\_width[:50], 'o', color='blue', label='setosa')

plt.plot(x\_petal\_length[50:], x\_petal\_width[50:], 'x', color='red', label='versicolor')

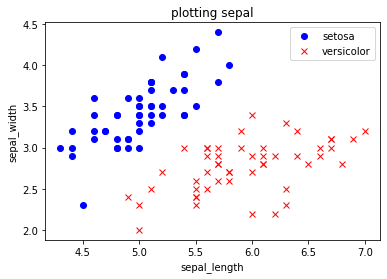
plt.title('plotting petal')

plt.xlabel('petal\_length')

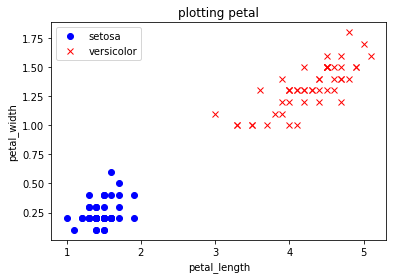
plt.ylabel('petal\_width')

plt.legend(); plt.show()

Dari kode di atas, mula-mula data diambil dari dataset setelah itu membagi data menjadi 2, yaitu input (x) dan output (y). Input berisi 4 parameter sehingga berisi matriks dimensi 1x4, sementara output berisi 1 output sehingga berbentuk matriks 1x1. Karena yang digunakan dalam tugas ini hanyalah 2 kelas output, maka hanya digunakan 100 data pertama. Kemudian, dari data input dibagi menjadi 4 data berdasarkan 4 parameter input. Untuk mendapatkan bayangan persebaran data, dilakukan plotting berdasarkan keempat parameter tersebut. Plotting dibagi menjadi 2, yaitu berdasarkan panjang & lebar petal dan berdasarkan panjang & lebar sepal. Diperoleh grafik seperti ganbar di bawah ini.

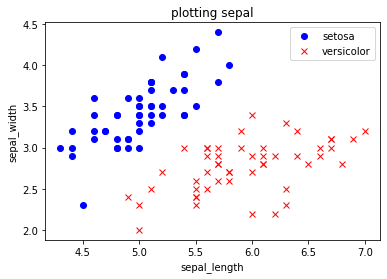


Gambar 1 *Scatter Plotting* Model Sepal

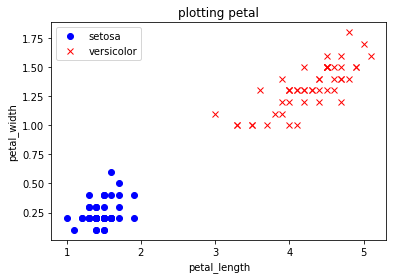


Gambar 2 *Scatter Plotting* Model Petal

Dari *plotting* di atas bisa diamati bahwa antara kelas *setosa* dan kelas *versicolor* terpisah secara ke dalam daerahnya masing-masing. Sehingga untuk melakukan klasifikasi dibutuhkan pemisah antara kedua daeraah tersebut. Untuk gambar 2, yaitu model berdasarkan data petal bisa kita lihat bahwa kelas *setosa* memiliki panjang petal kurang dari 2,5 dan lebar petal kurang dari 0,75 (angka diperoleh dari pengamatan *scatter plotting*). Batas ini bisa digunakan untuk melakukan klasifikasi dari kedua kelas. Jika kedua parameter tersebut kurang dari batas tersebut, maka akan diklasifikasikan sebagai kelas *setosa*, jika tidak, akan diklasifikasikan sebagai kelas *versicolor*. Sedangkan untuk gambar 1, yaitu model berdasarkan sepal bisa kita pisahkan antara kedua daerah kelas dengan sebuah garis miring. Untuk membuat garis dibutuhkan minimal 2 titik. Dari pengamatan *scatter plottnig* bisa diambil 2 titik yang kemungkinan menjadi garis pemisah antara kedua kelas. Diambil titik (5; 2,5) dan (6; 4). Dari kedua titik tersebut dibuat sebuah garis dengan parameter x adalah panjang sepal dan parameter y adalah lebar sepal. Diperoleh persamaan y = 1,5x -5. Nilai x disubstitusikan dengan nilai panjang sepal dan jika hasilnya lebih besar dari y yang disubstitusikan dengan nilai lebar sepal, maka akan diklasifikasikan ke dalam kelas *setosa*. Begitu pula sebaliknya, jika kurang dari maka diklasifikasikan sebagai *versicolor*.



Gambar 3 Ilustrasi Pembatas Plotting Sepal (Warna Hijau)



Gambar 4 Ilustrasi Pembatas Plotting Petal (Warna Hijau dan Kuning)

Dari kedua model di atas dihitung masing-masing akurasinya berdasarkan ketepatan dalam mengklasifikasikan kelas bunga *Irish*. Akurasi bisa dihitung dengan formula .

#predict based petal

def predict\_petal(x\_length, x\_width):

  if(x\_length<2.5 and x\_width<0.75):

    return 0

  else:

    return 1

#predict based sepal

def predict\_sepal(x\_length, x\_width):

  if((3/2\*x\_length - 5)<x\_width):

    return 0

  else:

    return 1

#counting accuration based petal

right\_predict = 0

for i in range(len(y)):

  if(predict\_petal(x\_petal\_length[i], x\_petal\_width[i]) == y[i]):

    right\_predict += 1

acc\_petal = right\_predict/len(y)\*100

print("Akurasi berdasarkan model pengamatan petal = " + str(acc\_petal) + "%")

#counting accuration based sepal

right\_predict = 0

for i in range(len(y)):

  if(predict\_sepal(x\_sepal\_length[i], x\_sepal\_width[i]) == y[i]):

    right\_predict += 1

acc\_sepal = right\_predict/len(y)\*100

print("Akurasi berdasarkan model pengamatan sepal = " + str(acc\_sepal) + "%")

Dari kode di atas, didapatkan output sebagai berikut,

Akurasi berdasarkan model pengamatan petal = 100.0%

Akurasi berdasarkan model pengamatan sepal = 99.0%

Dari hasil di atas, diperoleh bahwa untuk model berdasarkan pengamatan petal diperoleh akurasi 100% yang berarti semua hasil klasifikasi bunga *Irish* semuanya benar, tanpa ada kesalahan sedikitpun (dalam batasan dataset yang digunakan). Sementara, untuk model sepal diperoleh akurasi 99% yang berarti model bisa memprediksi 99 data dengan benar dan 1 data dengan hasil salah. Kesalahan tersebut bisa terjadi karena pembatas, garis miring, yang dibuat hanyalah berdasarkan pengamatan mata manusia dengan mengambil 2 titik yang paling representatif menurut pengamatan manusia. Hal tersebut tentunya akan menimbulkan *human error*. Oleh karena itu, model sepal tidak sebaik model petal, walaupun kedua akurasi dari kedua model tersebut bisa dikatakan cukup tinggi. Jadi, dalam mengklasifikasikan bunga *Irish* paling baik digunakan model berdasarkan pengamatan panjang dan lebar dari petalnya.