

Pembelajaran Mesin
Semester Genap Tahun Akademik 2017-2018
Tugas 1.3: Jaringan Saraf Probabilistik

Mochamad Yusuf Solihin
1301150020
IF 39-06

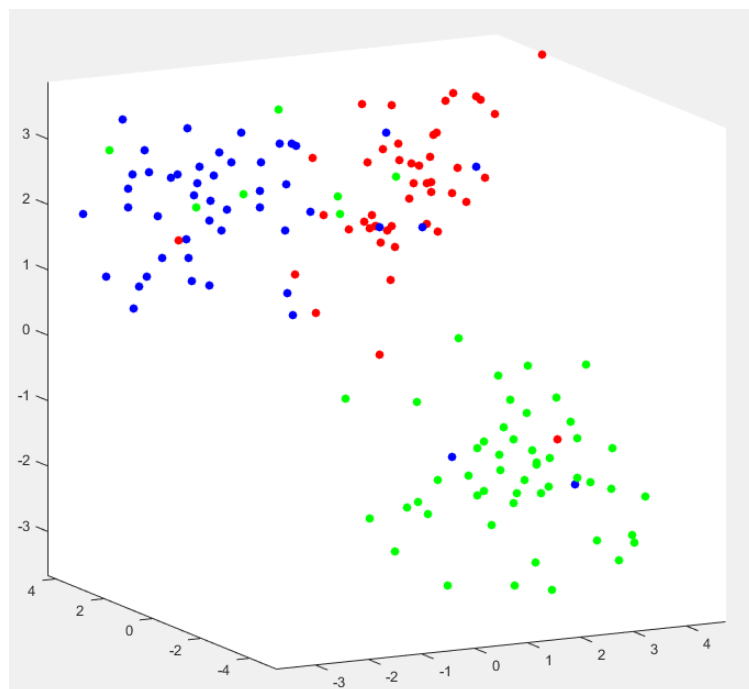
Februari 29, 2018

CLO3 (Total 100 poin)

A. Yang harus Anda lakukan saat proses pembangunan model:

1. [10 POIN] Load data latih dari file yang diberikan, visualisasikan seluruh data menggunakan scatter plot. Proses visualisasi ini dapat membantu Anda memahami persebaran data.

Jawab:



Dengan menggunakan scatter plot, persebaran data akan terlihat. Pada gambar diatas, dibagi menjadi 3 buah kelas yang masing-masing warna nya adalah merah untuk kelas 0, hijau untuk kelas 1, dan biru untuk kelas 2.

2. [25 POIN] Bangunlah fungsi-fungsi utama untuk mengklasifikasikan sebuah data menggunakan metode Jaringan Saraf Probabilistik. Pastikan fungsi-fungsi dengan jelas memperlihatkan setiap proses perhitungan di dalam Jaringan Saraf Probabilistik. Berikan baris komentar yang informatif pada kode jika diperlukan.

Jawab:

Fungsi mencari Smoothing Parameter

```

1 % fungsi mencari TAW untuk dimasukan ke fungsi PDF
2 function taw = fungsidelta(jmldata,kelasnya)
3 g = 1.5; % parameter untuk memperhalus hasilnya
4
5 % melakukan perulangan untuk mencari total jarak antar data dalam satu
6 % kelas yang sama yang nanti nya dicari jarak terpendeknya
7 for i = 1:jmldata
8     x = 1;
9     for j = 1:jmldata
10        if i ~= j
11            b(x) = norm([kelasnya(i,1) kelasnya(i,2) kelasnya(i,3)]-[kelasnya(j,1) kelasnya(j,2) kelasnya(j,3)]);
12            x = x+1;
13        end
14    end
15    jarak_terkecil(i) = min(b(:,1));
16 end
17 a = sum(jarak_terkecil);
18 z = a/jmldata;
19 taw = g*z/jmldata;
20 end

```

Fungsi PDF

```

1 % persamaan untuk mencari probabilitas PDF
2 function fx = fungsiPNN(taw,jmldataset,datatest1,datatest2,datatest3,datatrain1,datatrain2,datatrain3,jmldataset)
3 phi = 3.14;
4 m = 3; %m adalah dimensi nya
5 bawahkiri = ((2*phi).^(m/2))*(taw.^(m))*jmldataset;
6 bawahkanan = 2*(taw.^2);
7 x = 1;
8
9 %melakukan perulangan untuk mendapatkan hasil zigma (total dari
10 %exponensial nya
11 for i = 1:jmldataset
12     exponen(x,:) = exp(-(((datatest1 - datatrain1(i,1)).^2)+...
13         ((datatest2 - datatrain2(i,2)).^2)+...
14         ((datatest3 - datatrain3(i,3)).^2)))/bawahkanan;
15     x = x+1;
16 end
17 n_exponen = sum(exponen); %hasil penjumlahan exponen
18
19 fx = n_exponen/bawahkiri; %outputan
20 end

```

3. [30 POIN] Lakukan observasi untuk menentukan parameter-parameter terbaik yang akan digunakan di proses pengujian. Tuliskan laporan (Laporan [NIM] {KELAS}.pdf) mengenai detail observasi yang telah Anda lakukan. Tampilkan hasil visualisasi data dan grafik-grafik yang didapat saat observasi di laporan. Tuliskan juga hasil-hasil observasi yang digunakan oleh sistem pada saat pengujian.

Jawab:

Pada kasus kali ini digunakan metode PNN dengan rumus PDF untuk *single population*

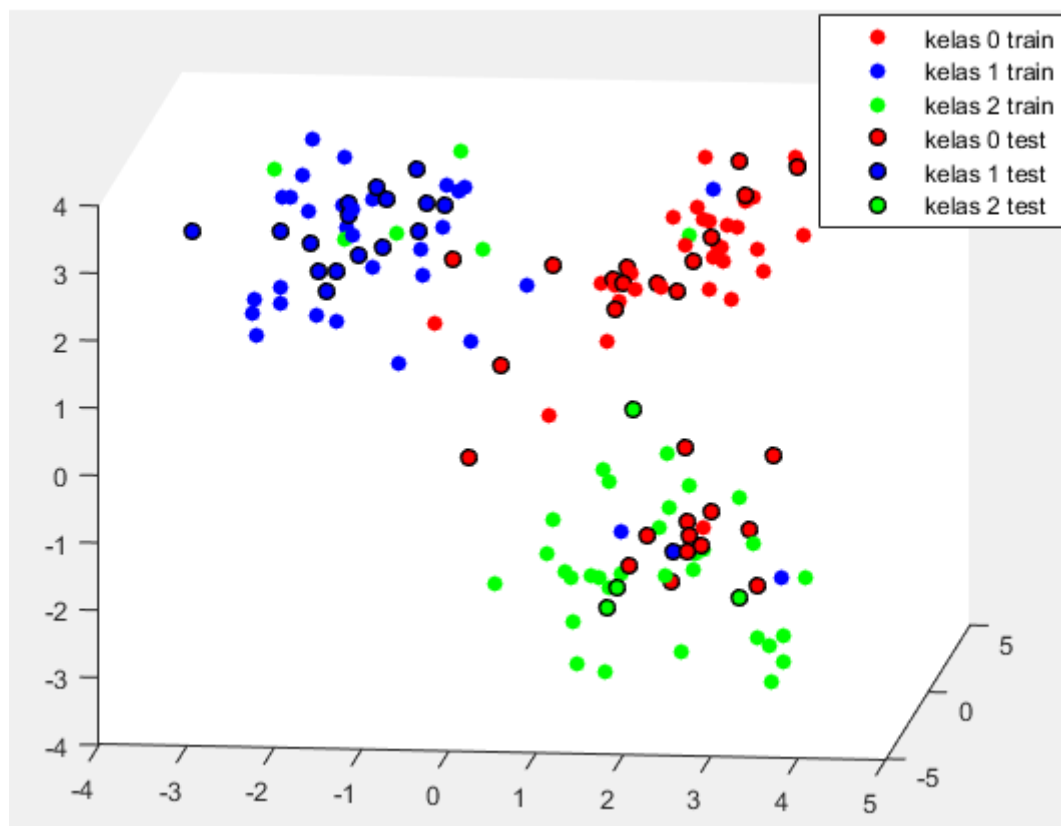
$$p(x|C_k) \approx \frac{1}{(2\pi)^{m/2} \sigma_k^m |C_k|} \sum_{\rho_i \in C_k} \exp\left[-\|x - w_i\|^2 / (2\sigma_k^2)\right]$$

Konstanta $\Phi = 3.14$ dan $m = 3$. Karena pada PDF diatas digunakan σ untuk menjadi *smoothing parameter*nya. Sedangkan untuk mencari σ adalah dengan menggunakan rumus :

$$\sigma_k = (g \cdot d_{tot}[k]) / |C_k|$$

$d_{tot}[k]$ adalah rata-rata jarak terdekat antar data untuk setiap kelasnya. kemudian $|C_k|$ adalah total data untuk setiap kelasnya. Dan g adalah parameter penghalus nya untuk menentukan nilai σ yang lebih akurat.

Dengan menggunakan rumus PNN diatas, maka dilakukan training untuk data_train yang dibagi menjadi dua bagian. 1-100 data untuk data train dan 101-150 untuk data train yang dijadikan data testnya. Tujuan nya adalah untuk mendapatkan tingkat akurasi program yang telah dibuat. Maka didapatkan hasil akurasi dari *training* yang dilakukan adalah 60% untuk 50 data train yang telah dijadikan data test.



Dari gambar pesebaran data diatas, didapatkan akurasi 60% dengan jumlah benar sebanyak 30 data, dan 20 nya bernilai salah.

B. Yang harus bisa dilakukan oleh sistem pada saat pengujian :

1. [25 POINT] Sistem Pengujian.

- a. Load data latih dan data uji dari file yang diberikan.

```
% -----  
% % import data_train dan data_test nya dari file excel  
% -----  
class = xlsread('data_train.xlsx','Sheet1','D1:D150');  
data_train = xlsread('data_train.xlsx','Sheet1','A1:D150');  
data_test = xlsread('data_test.xlsx','Sheet1','A1:C30');
```

- b. Lakukan proses klasifikasi terhadap data uji menggunakan metode Jaringan Saraf Probabilistik dengan parameter yang sudah Anda tentukan saat proses observasi.

```
% -----  
% % mencari nilai delta untuk setiap kelas  
% % dengan memanggil fungsi delta pada file yang lain  
% -----  
taw0 = fungsidelta(size(kelas0,1),kelas0);  
taw1 = fungsidelta(size(kelas1,1),kelas1);  
taw2 = fungsidelta(size(kelas2,1),kelas2);  
  
% -----  
% % menghitung probabilitas setiap kelas untuk dijadikan sigma  
% % yang nantinya bakal diambil kesimpulan untuk si data test  
% -----  
x = 1;  
for i = 1:size(data_test,1)  
    tabelhasil(i,1) = fungsiPNN(taw0,size(data_test,1),data_test(i,1),data_test(i,2),data_test(i,3),kelas0,kelas0,kelas0,size(kelas0,1));  
    tabelhasil(i,2) = fungsiPNN(taw1,size(data_test,1),data_test(i,1),data_test(i,2),data_test(i,3),kelas1,kelas1,kelas1,size(kelas1,1));  
    tabelhasil(i,3) = fungsiPNN(taw2,size(data_test,1),data_test(i,1),data_test(i,2),data_test(i,3),kelas2,kelas2,kelas2,size(kelas2,1));  
    x = x+1;  
end
```

```

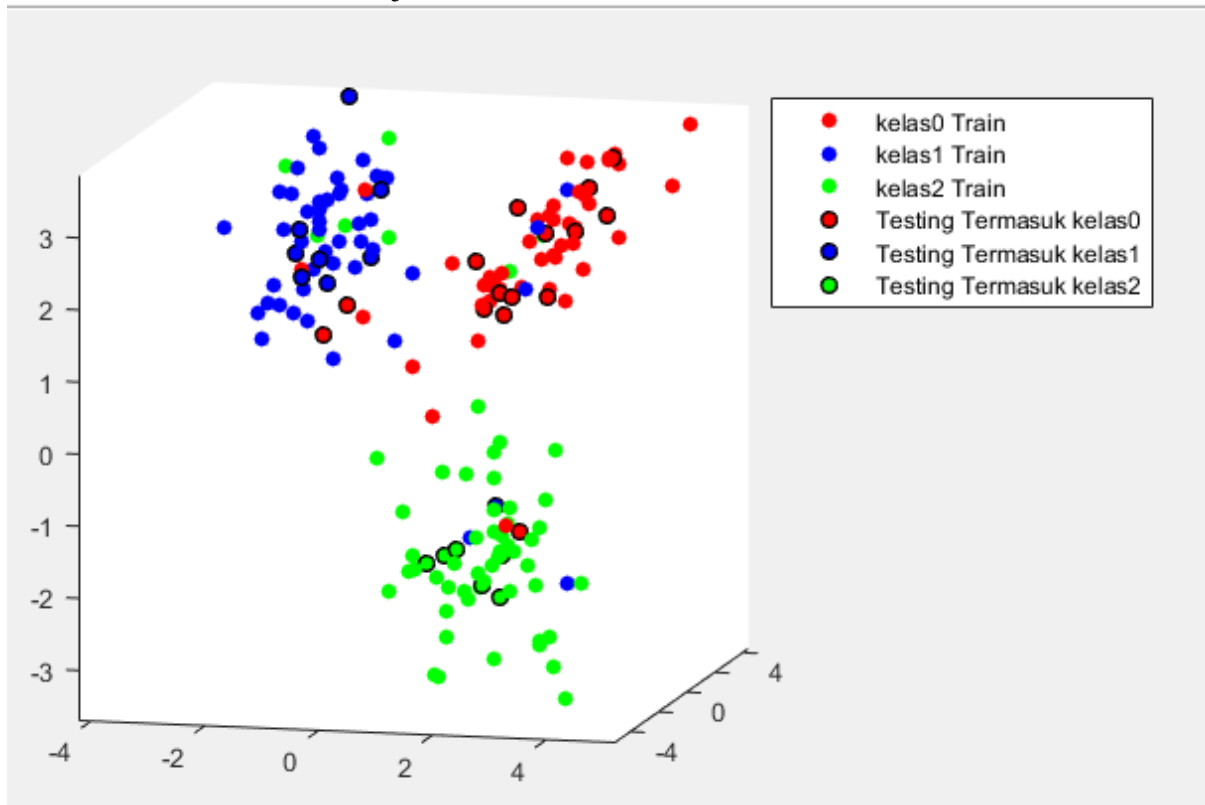
% -----
% % mencari nilai yang terbesar diantara data yg didapat
% -----
y = 1;
for i = 1:size(tabelhasil,1)
    finalkelas(y,:) = max(tabelhasil(i,:));
    y = y+1;
end

% -----
% % memasukan nilai probabilitas terbesar kedalam tabel hasil
% -----
tabelhasil(:,4) = finalkelas(:,:);

% -----
% % mengelompokan data data yang sudah didapat
% % kedalam kelas kelas yang bersangkutan
% -----
for i = 1:size(tabelhasil,1)
    if tabelhasil(i,4) == tabelhasil(i,1)
        tabelhasil(i,5) = 0;
    elseif tabelhasil(i,4) == tabelhasil(i,2)
        tabelhasil(i,5) = 1;
    elseif tabelhasil(i,4) == tabelhasil(i,3)
        tabelhasil(i,5) = 2;
    end
end
end

```

2. [10 POINT] Akurasi data uji



Dari grafik pesebaran data diatas dapat kita lihat bahwa warna merah untuk menunjukan pesebaran data train pada kelas 0, kemudian warna biru untuk menunjukan pesebaran data train pada kelas 1, warna hijau untuk menunjukan pesebaran data train pada kelas 2, warna merah dengan garis hitam untuk menunjukan pesebaran data test yang termasuk ke dalam kelas 0, warna biru dengan garis hitam untuk menunjukan pesebaran data test yang termasuk ke dalam kelas 1, warna hijau dengan garis hitam untuk menunjukan pesebaran data test yang termasuk ke dalam kelas 2.

Jika dilihat oleh mata manusia, maka didapatkan perkiraan kesalahan pada data test adalah sebanyak 4 titik data test karena 4 titik tersebut berada di daerah pesebaran data yang tidak sesuai warnanya. Contoh ada titik berwarna merah dengan garis hitam pada daerah data warna biru (kelas1), jika dilihat dari metode PNN dengan menggunakan *distance*, maka seharusnya titik data test tersebut berwarna biru dengan garis hitam. Oleh karena itu mungkin tingkat akurasi dari program yang dibuat kurang baik sehingga terjadi kesalahan klasifikasi pada beberapa data test.

References

- [1] Milo, "Probabilistic Neural Networks," *Probabilistic Neural Networks*, p. 58, 1 January 1990.
- [2] S. M. Dr. Suyanto, "Machine Learning," *ANN: Multi-Layer Perceptron & Probabilistic NN*, pp. 92 - 100, 2017.
- [3] V. Cheung, "An Introduction to," pp. 1-15, 10 june 2002.