

Statistické rozpoznávání - závěrečný projekt

David Müller

1 Analýza hlavních komponent - postup

1.1 Vycentrování dat

$$y_{ij} = x_{ij} - T$$

kde T je těžiště soustavy

$$T = \frac{1}{N} \sum_{i=1}^m \vec{x}_i$$

1.2 Vypočítání kovarianční matice

$$\text{cov}(X) = \frac{1}{N} \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T$$

1.3 Vlastní čísla a vlastní vektory

$$\det(A - \lambda E) = 0$$

1.4 Přepsání do jiné souřadné soustavy

$$Y = XW$$

kde W je matice obsahující na diagonále vlastní vektory

1.5 Whitening

$$y_k = \sum_{i=1}^n \frac{n_{k,i}}{\sqrt{\lambda_j}} x_j$$

2 Kód v Matlabu

```
1 close
2 clear
3 clc
4 %blood preassure
5 load iris_dataset.mat
6 X = x'
7
8 D = 2;
```

```

9  m=size(X,1);
10 t_x=mean(X,1);
11
12 %PCA
13 [Y,t,W,eff,eigvec_Y,eigval_Y,EigVecMat,EigValMat] = MY_PCA(X,D);
14
15 %whitening
16 Y_whitening = DATA_WHITENING_PCA(D,X)
17
18 %rozdělení do tříd pomocí DBscan
19 idx = dbscan(Y_whitening,0.48,D+1);
20 nout = sum(idx==-1)
21 nclass = max(idx)
22
23 set_index_outlayer = RETURN_INDEX_CLUSTER(Y_whitening,idx,-1);
24 set_index_1 = RETURN_INDEX_CLUSTER(Y_whitening,idx,1);
25 set_index_2 = RETURN_INDEX_CLUSTER(Y_whitening,idx,2);
26 set_index_3 = RETURN_INDEX_CLUSTER(Y_whitening,idx,3);
27
28 %Vykreslení grafu
29 figure(1)
30     subplot(2,2,1)
31         plot(X(:,1),X(:,2),"ko")
32         title("Input Data")
33     subplot(2,2,2)
34         plot(Y(:,1),Y(:,2),"ko")
35         title("Data after PCA")
36     subplot(2,2,3)
37         plot(Y_whitening(:,1),Y_whitening(:,2),"ko")
38         title("Data after whitening")
39     subplot(2,2,4)
40         plot(Y(:,1),Y(:,2),"ko")
41         hold on
42         plot([0 EigVecMat(1)*6],[0 EigVecMat(2)*6])
43         hold on
44         plot([0 EigVecMat(3)*6],[0 EigVecMat(4)*6])
45         hold off
46         legend("Data","1st component","2nd component")
47         title("1st and 2nd component")
48 figure(2)
49     plot(Y(set_index_1,1),Y(set_index_1,2),"r+")
50     hold on
51     plot(Y(set_index_2,1),Y(set_index_2,2),"b+")
52     hold on
53     plot(Y(set_index_3,1),Y(set_index_3,2),"m+")

```

```

54         hold on
55         plot(Y(set_index_outlayer,1),Y(set_index_outlayer,2),"ko")
56         hold off
57         legend("class 1","class 2","class 3","outlayer")
58         title("Clustering")
59

```

2.1 Funkce - MY_PCA

```

1  function [Y,t,W,eff,eigvec_Y,eigval_Y,EigVecMat,EigValMat]= MY_PCA(X,D)
2  %Y...komponenty
3  %t...teziste
4  %W...matice vektoru vah komponent - faktory
5  %eff...efektivita
6  %X...data
7  %D...pocet komponent
8
9  m=size(X,1);
10 t=mean(X,1);
11
12 % 1) vycentrovani X
13 X=X-repmat(t,m,1);
14 t=t';
15
16 % 2) kovariancni matice
17 CovMat=X'*X/(m-1);
18
19 % 3) vlastni cisla (L) a vektory (V)
20 [EigVecMat,EigValMat]=eig(CovMat);
21 lambda = diag(EigValMat);
22 % 3a) serazeni
23 [lambda,ind]=sort(lambda,"descend");
24 EigVecMat=EigVecMat(:,ind);
25 s=cumsum(lambda);
26
27 % 4) efektivita
28 eff = s(1:D)/s(end);
29
30 % 5) PCA
31 EigVecMat=EigVecMat(:,1:D);
32 W=EigVecMat';
33
34 Y=X*W';
35 m_Y=size(Y,1);

```

```

36 C_Y=Y'*Y/(m_Y-1);
37 [eigvec_Y,eigval_Y]=eig(C_Y);
38 end
39

```

2.2 Data Whitening

```

1
2 for i = 1:D
3     W(i,:)=W(i,+)/sqrt(lambda(i));
4 end
5 Y=X*W';
6 end

```

2.3 Funkce pro vizualizaci clusteringu

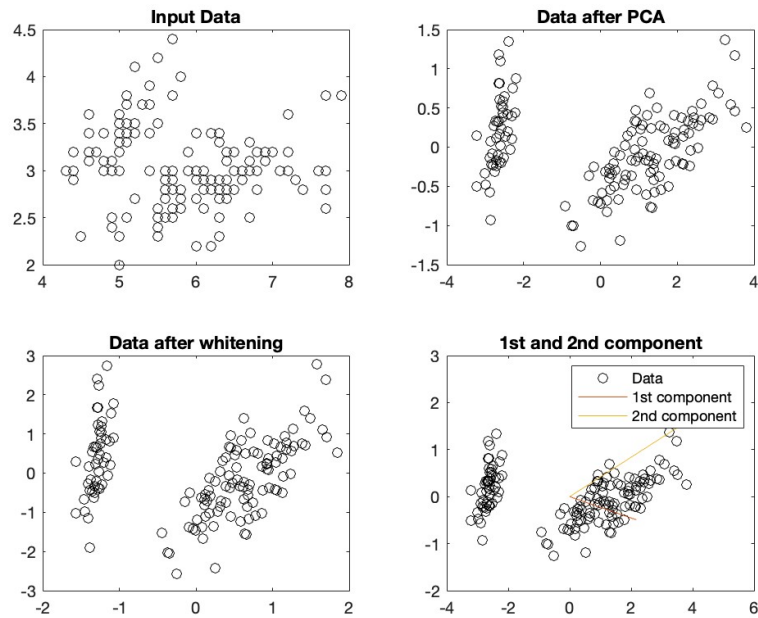
```

1
2 function [set_index] = RETURN_INDEX_CLUSTER(Y,idx,class)
3 set_index = [];
4     for i =1:length(Y)
5         if idx(i) == class
6             set_index=[set_index;i];
7         end
8     end
9
10 end

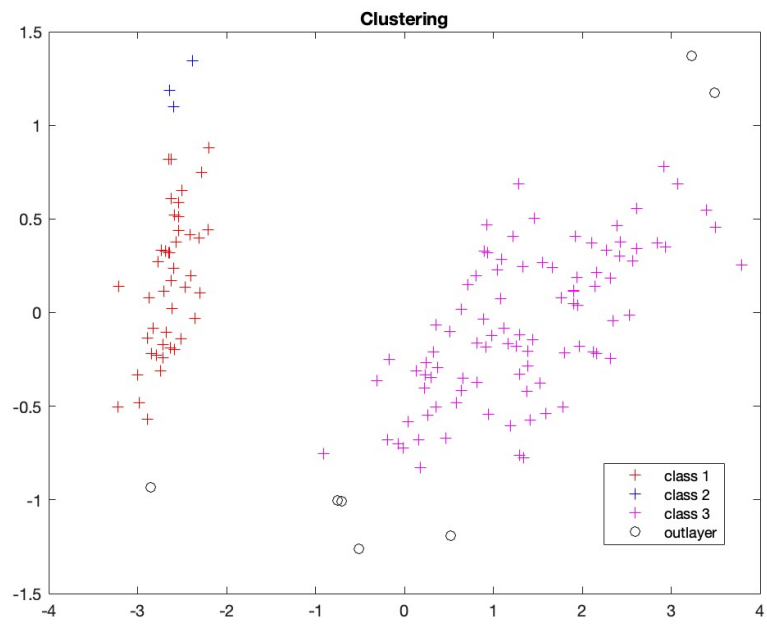
```

2.4 Výstup

Zobrazení původních dat a dat po PCA a Data Whiteningu



Byl proveden clustering pomocí funkce "dbscan" s parametry "epsilon" (poloměr pro hledání sousedů) 0.48 a "minpts" (minimální počet sousedů) 3



Data se rozdělily do 3 tříd a vzniklo 7 outlayerů.