Homework Set 3, CPSC 8420, Fall 2020

Last Name, First Name

Due 10/29/2020, Thursday, 11:59PM EST

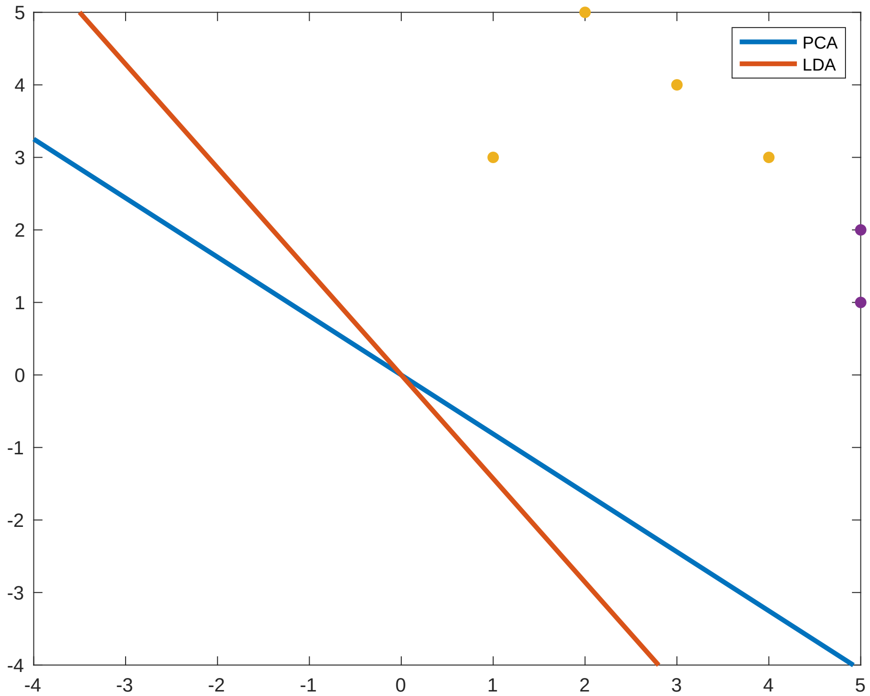
Problem 1

Given datapoints $\{\{1,3\},\{2,5\},\{3,4\},\{4,3\},\{5,2\},\{5,1\}\}$.

1. Please scatter-plot each datapoint within one figure (you can use Matlab, Python or any other programming language).
2. Now if we want to reduce the dimension from 2 to 1 by PCA, please determine the projection line which crosses the origin (please plot the line based on the scatter-plot figure above).
3. Assume the first 4 datapoints belong to one class, while the rest 2 belong to the other. Now if we want to reduce the dimension from 2 to 1 by LDA, please determine the projection line which crosses the origin (you are expected to plot the line based on the scatter-plot figure).

```
close all;
a = [1,3;2 5;3 4;4 3];
b=[5 2;5 1];
c= [a;b];
c_bar=c-mean(c);
[W_pca,~,~]=svd(c_bar'*c_bar);
d=a-mean(a);
e=b-mean(b);
Sw=d'*d+e'*e;
W_lda=Sw\(mean(a)-mean(b))';
x=-4:0.1:5;
y=(W_pca(2)/W_pca(1)).*x;
figure
plot(x,y,'LineWidth',2.5)
xlim([-4,5]);
ylim([-4,5]);
hold on;
z=(W_lda(2)/W_lda(1)).*x;
plot(x,z,'LineWidth',2.5)
hold on;
x=a(:,1);
y=a(:,2);
scatter(x,y,'filled')
hold on;
```

```
x=b(:,1);
y=b(:,2);
scatter(x,y,'filled')
legend('PCA','LDA')
```



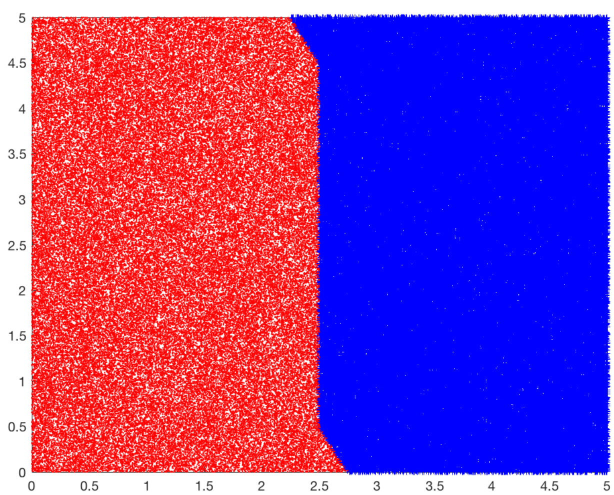
Problem 2

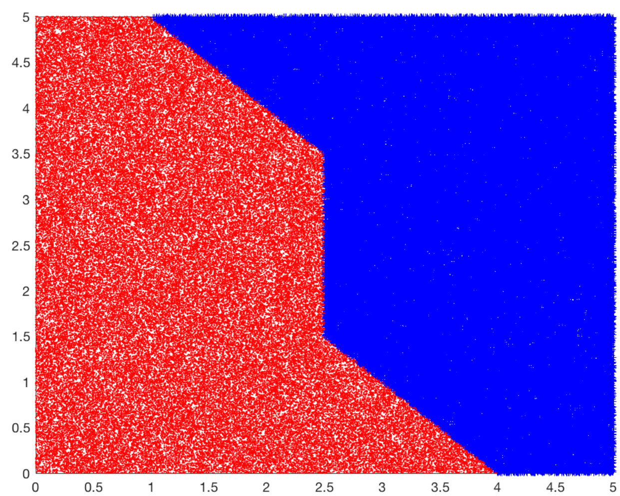
Given positive dataset $\{\{1,1\},\{2,2\},\{2,3\}\}$, as well as negative dataset $\{\{3,2\},\{3,3\},\{4,4\}\}$, please determine the decision boundary when leveraging k-NN where $k=1$ and $k=3$ respectively (should be similar to the figures on slide 5–6 of Lecture 2).

```
clear all;close all;
data_size=50000;%you may set it smaller say 50000
a = 5*rand(data_size,2);
label=zeros(data_size,1);
distance_all=zeros(data_size,6);
for i=1:size(a,1)
    distance_all(i,1)=(a(i,1)-1)^2+(a(i,2)-1)^2;
    distance_all(i,2)=(a(i,1)-2)^2+(a(i,2)-2)^2;
    distance_all(i,3)=(a(i,1)-2)^2+(a(i,2)-3)^2;
    distance_all(i,4)=(a(i,1)-3)^2+(a(i,2)-2)^2;
    distance_all(i,5)=(a(i,1)-3)^2+(a(i,2)-3)^2;
    distance_all(i,6)=(a(i,1)-4)^2+(a(i,2)-4)^2;
end
[~,mask] = myMaxk(50-distance_all',1);%3 denotes the K in KNN, you may change it to 1
for i=1:size(a,1)
```

```
    c = mask(1:3,i);
    b = mask(4:6,i);
    if sum(c)>sum(b)
        plot(a(i,1),a(i,2),'r.','MarkerSize',3);hold on;
    else
        plot(a(i,1),a(i,2),'b+','MarkerSize',3);hold on;
    end
end

function [ C,mask_1 ] = myMaxk( A,b )
    signs=sign(A);
    AA=abs(A);
    [~,idx]=sort(AA,1);
    idx_mask=idx(1+size(idx,1)-b:end,:);
    mask_1=zeros(size(A));
    for j=1:size(idx_mask,2)
        for i=1:size(idx_mask,1)
            mask_1(idx_mask(i,j),j)=1;
        end
    end
    C=AA.*signs.*mask_1;
end
```





Problem 3

Considering soft margin SVM, where we have the objective and constraints as follows:

$$
\begin{aligned}
& min \;\; \frac{1}{2}||w||_2^2 + C\sum_{i=1}^{m}\xi_i \\
& s.t. \;\; y_i(w^T x_i + b) \geq 1 - \xi_i \;\; (i = 1, 2, ...m) \\
& \xi_i \geq 0 \;\; (i = 1, 2, ...m)
\end{aligned} \tag{1}
$$

Now we formulate another formulation as:

$$\begin{aligned} &min \ \ \frac{1}{2}||w||_2^2 + \frac{C}{2}\sum_{i=1}^{m}\xi_i^2 \\ &s.t. \ \ y_i(w^Tx_i{+}b) \geq 1 - \xi_i \ \ (i = 1, 2, ...m) \end{aligned} \tag{2}$$

1. Different from Eq. (1), we now drop the non-negative constraint for ξ_i, please show that optimal value of the objective will be the same when ξ_i constraint is removed.
   ξ_i can't be negative, that is assume it is negative, then the constraint $y_i(w^Tx_i + b) \geq 1 - \xi_i$ also satisifies for $\xi_i = 0$ while the objective function would be lower.

2. What's the generized Lagrangian of the new soft margin SVM optimization problem?
   $\frac{1}{2}||w||_2^2 + \frac{C}{2}\sum\limits_{i=1}^{m}\xi_i^2 - \sum\limits_{i=1}^{m}\alpha_i[y_i(w^Tx_i + b) - 1 + \xi_i]$, *where* $\alpha_i \geq 0 \ (i = 1, 2, ...m)$.

3. Now please minimize the Lagrangian with respect to w, b, and ξ.
   $w = \sum\limits_{i=1}^{m}\alpha_i y_i x_i, \ \sum\limits_{i=1}^{m}\alpha_i y_i = 0, \ C\xi_i = \alpha_i \ (i = 1, 2, ...m)$

4. What is the dual of this version soft margin SVM optimization problem? (should be similar to Eq. (10) in the slides)

$$\begin{aligned} &\underbrace{min}_{\alpha}\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j\langle x_i, x_j\rangle - \sum_{i=1}^{m}\alpha_i + \frac{1}{2C}\sum_{i=1}^{m}\alpha_i^2 \\ &s.t. \ \sum_{i=1}^{m}\alpha_i y_i = 0 \\ &\alpha_i \geq 0 \ i = 1, 2, ...m \end{aligned} \tag{3}$$

5. Please analysis bias-variance tradeoff when C increases.
   When C increases, the bias decreases and variance increases since it is less tolerant with misclassification, thus the margin decreases, the generalization ability decreases.