Αναγνώριση Προτύπων

Εργασία 3

Αλβανάκη Παρασκευή

ΑΜ 57286

ΕΡΩΤΗΜΑ 3.1

A)

Ο κώδικας με τον οποίο υλοποιήθηκε η g(x) είναι ο παρακάτω:

```
function [g] = g_calculation(m,S,P,x)
d=length(m);
g=-0.5*(transpose(x-m))*(S^(-1))*(x-m)-d*0.5*log(2*pi)-0.5*log(det(S))+log(P);
end
```

B)

Ο κώδικας με τον οποίο υπολογίστηκε η Ευκλείδεια απόσταση μεταξύ δύο αυθαίρετων σημείων d-διαστάσεων x1 και x2είναι ο παρακάτω:

```
function [dist] = Euclidean(x1,x2)
dist=sqrt(sum((x1-x2).^2));
disp(dist);
%ή με norm(x1-x2)
end
```

Γ)

Η απόσταση Mahalanobis είναι ένα μέτρο της απόστασης μεταξύ ενός σημείου P(διανύσματος παρατηρήσεων) και μιας κατανομής D. Είναι μία πολυδιάστατη γενίκευση του πόσες τυπικές αποκλίσεις απέχει το P από τη μέση τιμή της κατανομής.

Ο κώδικας με τον οποίο υπολογίστηκε η Ευκλείδεια απόσταση μεταξύ δύο αυθαίρετων σημείων d-διαστάσεων x1 και x2είναι ο παρακάτω:

```
function [mah_dist] = Mahalanobis(m,S,x)
mah_dist=sqrt((transpose(x-m)*inv(S)*(x-m));
end
```

ΕΡΩΤΗΜΑ 3.2

1

Το σημείο διαχωρισμού για τις 2 κατηγορίες είναι $-4.8438<x1<4.0958$ για να έχει ταξινομηθεί στο ω1.

2

Τα εμπειρικά σφάλματα για τα δείγματα που μας δώθηκαν είναι:

e1= -4.8438 και e2=4.0958

3

Η περιοχή διαχωρισμού για τα 2 χαρακτηριστικά είναι η παρακάτω:
- 0.023547938*x1^2 - 0.017614759*x1 + 0.004196469*x2^2 - 0.06859997*x2 +0.323072535
και τα σφάλματα είναι e1=0.2 , e2=0.2

4

Η περιοχή διαχωρισμού για τα 3 χαρακτηριστικά είναι η παρακάτω:
- 0.023547938*x1^2 - 0.017614759*x1 + 0.004196469*x2^2 - 0.0685999705*x2 + 0.0019927526*x3^2 - 0.010458093*x3 + 0.2812254608
και τα σφάλματα είναι e1=0.2 , e2=0.2

5

Η προσθήκη περισσότερων χαρακτηριστικών δεν πετυχαίνει απαραίτητα καλύτερο διαχωρισμό . Αν αυξήσουμε το πλήθος των δειγμάτων του training τότε ίσως έχουμε καλύτερα αποτελέσματα με τα περισσότερα χαρακτηριστικά αν το range τον αποτελεσμάτων μας βοηθάει στο classification (αν δηλαδή τα δείγματα διαφορετικών κλάσεων ενός νέου χαρακτηριστικού απέχουν επαρκώς ως προς τον άξονα του).

6

Οι περιοχές διαχωρισμού είναι οι εξής:

D12=- 0.023547938*x1^2 - 0.017614759*x1 + 0.0041964693*x2^2 - 0.06859997*x2 + 0.0019927526*x3^2 - 0.01045809*x3 + 2.360667

D13=0.02826353466*x1^2 - 0.5535114818*x1 + 0.0269063612*x2^2 - 0.311493715*x2 - 0.016417189*x3^2 - 0.08027517969*x3 + 2.99500356869
D23=0.0518114727*x1^2 - 0.5358967227*x1 + 0.0227098919*x2^2 - 0.2428937447*x2 - 0.018409942*x3^2 - 0.0698170866*x3 + 0.634336566

ΕΡΩΤΗΜΑ 3.3

1

Η αρχική κατανομή του θ αποτελεί συνάρτηση πυκνότητας πιθανότητας και άρα το ολοκλήρωμα της ως προς θ είναι 1.
Ολοκληρώνοντας ως προς θ μπορουμε να βρουμε το Α.

```
p=sin(pi*x);
A=1/int(p,[0 1]);

>>A=pi/2
```

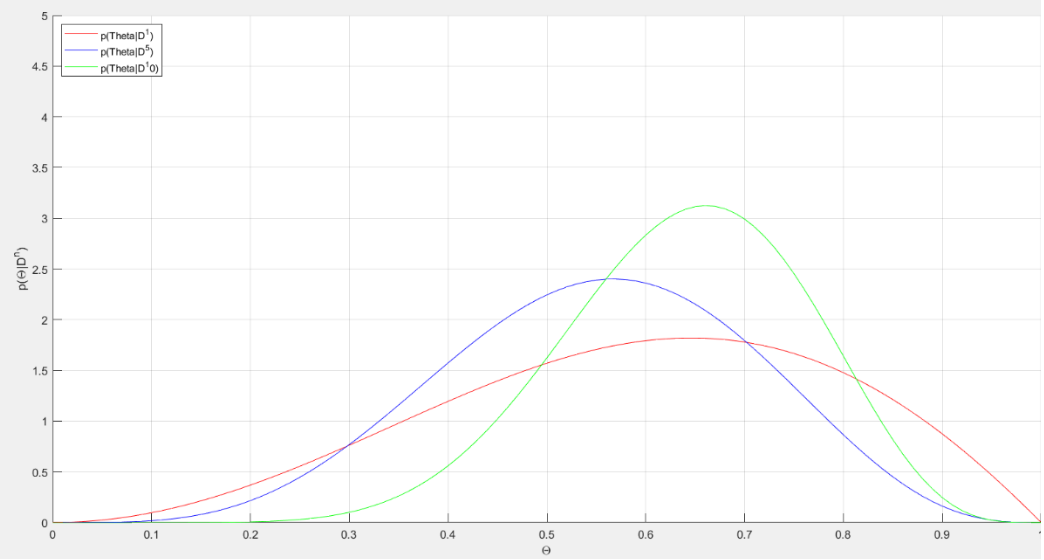
2

Γνωρίζω πως στο πρόβλημα ρίψης νομίσματος ισχύει:

$p(x_n|\theta)$* $p(\theta|D^{n-1})$=$\theta^{\kappa}*(1-\theta)^{n-\kappa}p(\theta|D^0)$

Συνεπώς προκύπτει η σχέση: $p(\theta|D^n)$=$\frac{\theta^{\kappa}*(1-\theta)^{n-\kappa}p(\theta|D^0)}{\int \theta^{\kappa}*(1-\theta)^{n-\kappa}p(\theta|D^0)d\theta}$

Η γραφική αποικόνιση είναι η παρακάτω:



3

Το p(x=k|D^10=3.1244 για χ=0.66 (Ο κώδικας με τον οποίο υπολογίστηκε βρίσκεται στο αρχείο ask_3)

ΕΡΩΤΗΜΑ 4

A

Δημιουργω τυχαια δείγματα για το X χρησιμοποιώντας τη συναρτηση ler1=mvnrnd(m1,s,3334) για την $1^{η}$ κατανομη και ler2=mvnrnd(m1,s,3333). Για την $2^{η}$ και $3^{η}$ (αυτές έχουν 1 τιμή λιγότερη αφού θέλω τα συνολικά σημεία να είναι 10000.Ομοίως δημιουργώ τα δειγματα για το X1 με την εντολη t1=mvnrnd(m1,s,334) Και t2=mvnrnd(m1,s,333) για τις t2 και t3.

B

Για το ερώτημα αυτό επέλεξα τον ταξινομιτή ο οποίος χρησιμοποιεί την απόσταση Mahalabobis που σε αυτή την περίπτωση είναι ισοδύναμος με τον Bayes.Ευκλείδια απόσταση δεν ενδείκνυται να χρησιμοποιήσω γιατι ο πίνακας Σ δεν είναι διαγώνιος. Στο αρχείο ask_4_b έχουν υλοποιηθεί και οι 3 ταξινομητές(με ευκλείδια με mahalanobis και με bayes και παρατηρούμε πως όντως ο Bayes δίνει ίδιο αποτέλεσμα με τον Mahalanobis και η ευκλείδια απόσταση δίνει το χειρότερο αποτέλεσμα όπως το περιμέναμε.

```
>> ask_4_b
The Mahalanobis error is:
    0.1000

Euclidean Error
    0.1100

The Bayes error is:
    0.1000
```

Ο κώδικας ο οποίος υπολογίζει την Mahalanobis distance και το αντίστοιχ error δινεται παρακάτω.Ουσιαστικά για κάθε δείγμα το ταξινομώ στην κλάση με την μικρότερη αποσταση και στη συνέχεια βλέπω εάν τα δείγματα μου εχουν ταξινομηθεί στη σωστή κατανομή.Ο κώδικας είναι ο εξης:

```
X1=[t1;t2;t3]';

[l,c]=size(m);
[l,N]=size(X1);
for i=1:N
    for j=1:c
        dm(j)=sqrt((X1(:,i)-m(:,j))'*s^-1*(X1(:,i)-m(:,j)));
    end
    [num,z(i)]=min(dm);

end

w1=(length(find(z(1:334)==1)));
```

```
w2=(length(find(z(335:667)==2)));
w3=(length(find(z(668:1000)==3)));
error=1-(w1+w2+w3)/1000;
disp("The Mahalanobis error is:");
disp(error);
```

Στο αρχείο περιλαμβάνονται και οι κώδικες για την ευκλείδια απόσταση και τον Bayes.

Γ

Για το ερώτημα αυτό χρειάζεται να υπολογίσουμε και τις πραμέτρους μ και Σ.Έτσι υπολογίζουμε για κάθε μια κατανομή τα μ1,μ2,μ3 και τα Σ1 Σ2 Σ3 και στη συνέχεια για να βγάλουμε το τελικό Σ υπολογίζουμε τον μέσο όρο των τριών Σ.Πάλι επιλέγεται η Mahalanobis distance για τους λόγους που αναφέρθηκαν και στο προηγούμενο ερώτημα ομοίως με πριν όμως υλοποιούνται και οι υπόλοιποι ταξινομιτές για να συγκρίνουμε τα αποτελέσματα.Παρατηρούμε ότι όντως και σε αυτό το ερώτημα οι ταξινομιτές bayes και mahalanobis δίνουν το ίδιο αποτέλεσμα με την ευκλείδια απόσταση να μας δίνει το μεγαλύτερο σφάλμα όπως αναμενόταν.

```
The Mahalanobis error is:
    0.0950

Euclidean Error
    0.0920

Bayesian Error
    0.0950
```

Ο κώδικας για τον ταξινομιτή mahalanobis είναι ο παρακάτω(στο αρχείο ask_4_c υπάρχουν υλοποιημένοι mahalanobis και eucleidian) :

```
%------------------ c Mahalanobis ----------
[l,c]=size(m);
[l,N]=size(X1);
x1=X(:,1:3334);
x2=X(:,3335:6667);
x3=X(:,6668:10000);
m_hat1=(1/length(x1)).*sum(x1,2);
m_hat2=(1/length(x2)).*sum(x2,2);
m_hat3=(1/length(x3)).*sum(x3,2);
S1=0;S2=0;S3=0;
for k=1:length(x2)
    S1=S1 + (x1(:,k)-m_hat1)*(x1(:,k)-m_hat1)';
    S2=S2 + (x2(:,k)-m_hat2)*(x2(:,k)-m_hat2)';
    S3=S3 + (x3(:,k)-m_hat3)*(x3(:,k)-m_hat3)';
end
S1=S1+(x1(:,length(x1))-m_hat1)*(x1(:,length(x1))-m_hat1)';
```

```
S_hat=(1/3)*((1/length(x1))*S1+(1/length(x2))*S2+(1/length(x3))*S3)
';
%κατηγοριοποίηση των test
m=[m_hat1 m_hat2 m_hat3];
[l,c]=size(m);
[l,N]=size(X1);
for i=1:N
    for j=1:c
        dm(j)=sqrt((X1(:,i)-m(:,j))'*s^-1*(X1(:,i)-m(:,j)));
    end
    [num,z(i)]=min(dm);

end


w1=(length(find(z(1:334)==1)));
w2=(length(find(z(335:667)==2)));
w3=(length(find(z(668:1000)==3)));
error2=1-(w1+w2+w3)/1000;
disp("The Mahalanobis error is:");
disp(error2);
```

Δ

Με τα συγκεκριμένα Σ που μας δίνονται δεν μπορεί να πραγματοποιηθεί ούτε mahalanobis ούτε η ευκλείδια απόσταση ως classifiers καθώς ο covariance matrix δεν είναι ίδιος για όλες τις κλάσεις.Συνεπώς θα χρησιμοποιήσω Bayes.O Bayes ουσιαστικά ταξινομεί κάθε δείγμα στην κλαση με το μεγαλύτερο $P(w_1|x) = \frac{p(x|w_1)\cdot P(\omega_1)}{p(x)}$. Για λόγους σύγκρισης όπως και πρίν έχουν δηλιουργηθεί και οι eyclidian και mahalanobis ταξινομιτές για τ αερωτήματα β και γ με τα νέα δεδομένα στο αρχείο ask_4_d.

- Για το ερώτημα β παρατηρούμε ότι όντως ο bayes μας δίνει το μικροτερο σφάλμα όπως αναμέναμε.

```
The Bayesian error for new b is:
    0.0380

The Mahalanobis error for new b is:
    0.0470

The Euclidean Error for new b is:
    0.0450
```

Ο κώδικας με τον οποίο υλοποιήθηκε ο Bayes είναι ο παρακάτω:

```
%bayesian
[l,c]=size(m);
[l,N]=size(X1);
```

```
lik=@(x,m,s,d) exp(-0.5*(x-m)'*inv(s)*(x-
m))/(((2*pi)^(d/2))*det(s)^(0.5));
z=zeros(1,length(X1));
for i=1:N
    dm=[p1*lik(X1(:,i),m1,s(:,:,1),3)
p2*lik(X1(:,i),m2,s(:,:,2),3) p3*lik(X1(:,i),m3,s(:,:,3),3)];
    z(i)=find(dm==max(dm));
end
w1=(length(find(z(1:167)==1)));
w2=(length(find(z(168:334)==2)));
w3=(length(find(z(335:1000)==3)));
error=1-(w1+w2+w3)/1000;
disp("The Bayesian error for new b is:");
disp(error);
```

- Για το ερώτημα γ παρατηρούμε ότι όπως και πριν ο bayes μας δίνει το μικροτερο σφάλμα όπως αναμέναμε.

```
The Bayesian error for new c is:
    0.0360

The Mahalanobis error for new c is:
    0.0460

The Euclidean Error for new c is:
    0.0440
```