

بسمه تعالی

پروژه احتمال مهندسی – پیاده سازی الگوریتم یادگیری *Naive Bayes*

مرتضی کریمی 40008423

مصحح گرامی: توجه فرمایید که در محاسبه توزیع *Bark Days* پس از دسته بندی *Bark Days* در هر کلاس، واریانس داده *Bark Days* در هر کلاس بیشتر از واریانس آن شد، لذا نمی توان به آن توزیع دوجمله ای به آن تخصیص داد، اما مطابق راهنمایی حل تمرین مقدار n را به طور پیش فرض برای هر سه کلاس 30 در نظر گرفتیم و سپس از روی میانگین و طبق رابطه $np = mean$ مقدار p را محاسبه کردیم.

شبیه سازی های توزیع های قسمت اول در یک *m-file* با عنوان *simulation1* و کد مشخص کننده دسته توزیع قسمت اول در یک *m-file* با عنوان *alg1* در ذخیره شده است.

ابتدا مروری بر ریاضیات و توضیحات مربوطه می کنیم؛

مطابق توضیحات داده شده در فایل پروژه، اگر که $X = (X_1, \dots, X_n)$ بردار ویژگی یک نمونه باشد و C_k به ازای $k = 1, \dots, m$ را یک دسته یا کلاس در نظر بگیریم، برای تعیین تعلق بردار X به دسته مورد نظر بایستی که احتمال $P(C_k|X)$ را محاسبه کنیم، و از میان احتمالات موجود مقدار بیشینه $P(C_k|X)$ دسته مورد نظر را مشخص خواهد کرد، از سوی دیگر توجه فرمایید که برای محاسبه احتمال مذکور می توان از رابطه بیز به ترتیب زیر استفاده کرد:

$$P(C_k|X) = \frac{P(X|C_k)P(C_k)}{P(X)}$$

اما بیایید احتمال $P(C_k|X)$ را سراسر تر بنویسیم، یعنی مستقیماً با استفاده از رابطه احتمال شرطی:

$$P(C_k|X) = \frac{P(C_k, X)}{P(X)}$$

توجه فرمایید که برای مشخص کردن کلاس میبایست که مقدار \max احتمال $P(C_k|X)$ را در رابطه فوق محاسبه نماییم، به ازای k ای که مقدار $P(C_k|X)$ بیشینه شود، C_k کلاس مورد نظر خواهد بود، $P(X)$ در تمامی $P(C_k)$ ها یک مقدار ثابت دارد، لذا بایستی که $P(C_k, X)$ را بیشینه کنیم؛ توجه فرمایید که احتمال $P(C_k, X)$ را با توجه به اینکه $X = (x_1, \dots, x_n)$ است را به صورت زیر می توان بازنویسی کرد:

$$P(C_k, X) = P(C_k, x_1, \dots, x_n) = P(x_1, \dots, x_n, C_k)$$

حال اگر که رابطه فوق را با توجه به قاعده زنجیری بازنویسی کنیم داریم:

$$\begin{aligned} P(C_k, X) &= P(C_k, x_1, \dots, x_n) = P(x_1, \dots, x_n, C_k) \\ &= P(x_1 | x_2, \dots, x_n, C_k) P(x_2 | x_3, \dots, x_n, C_k) \dots P(x_{n-1} | x_n, C_k) P(x_n | C_k) P(C_k) \end{aligned}$$

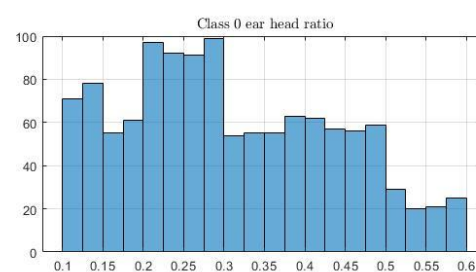
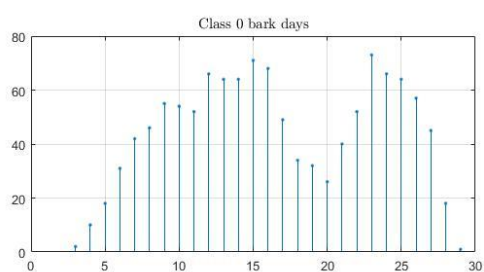
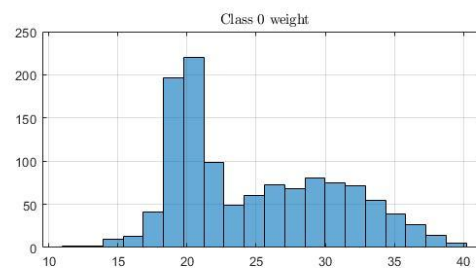
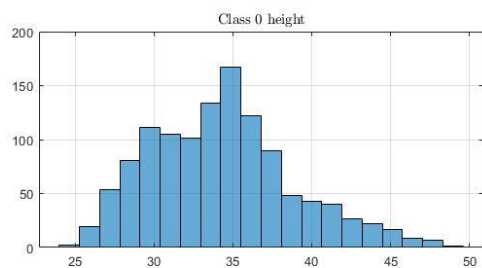
حال با فرض آنکه ویژگی ها در بردار ویژگی از یکدیگر استقلال آماری داشته باشند، می توان نوشت:

$$P(C_k, X) = P(x_1 | C_k) P(x_2 | C_k) \dots P(x_{n-1} | C_k) P(x_n | C_k) P(C_k)$$

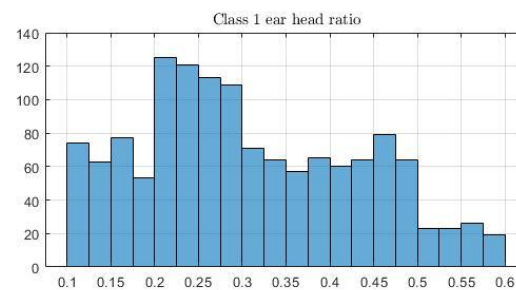
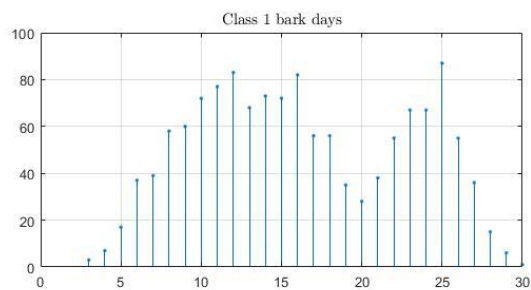
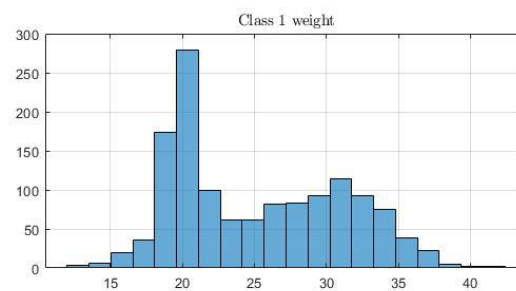
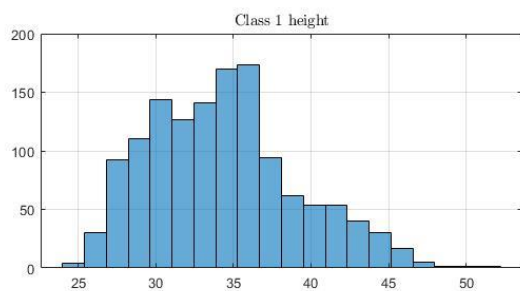
در هر دسته k ام که مقدار فوق ماکزیمم شود، آن دسته را انتخاب خواهیم کرد.

توزیع های بدست آمده برای هر ویژگی به تفکیک کلاس به صورت زیر است:

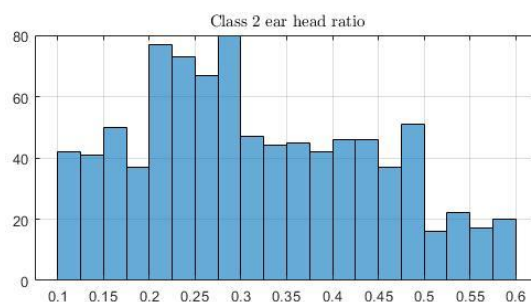
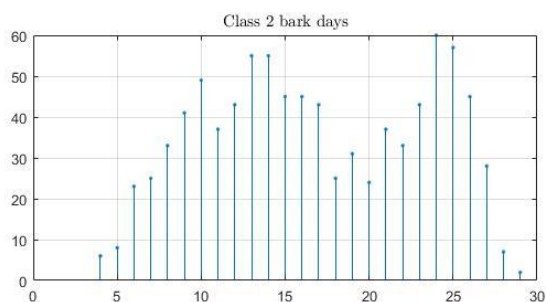
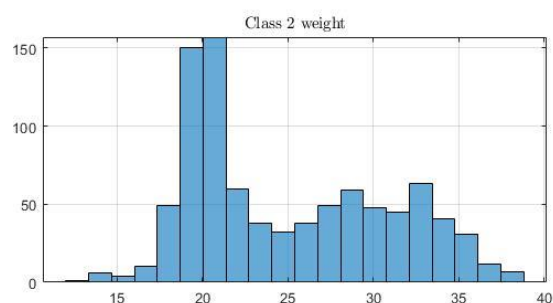
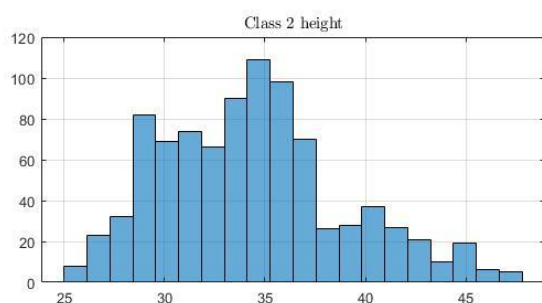
برای کلاس 0:



برای کلاس 1



برای کلاس 2:



در ماتریس PM مقادیر واریانس و میانگین های هر ویژگی را مشخص می کند، هر سطر مشخص کننده یک کلاس خاص است، در هر سطر، درایه فرد میانگین و درایه زوج واریانس را مشخص می کند، دو درایه اول هر سطر مربوط به ویژگی $height$ ، دو درایه بعدی مربوط به ویژگی $weight$ دو درایه بعدی مربوط به ویژگی $bark$ و دو درایه بعدی مشخص کننده ویژگی $ear\ head\ ratio$ هستند.

$PM =$

34.3038	20.9177	24.9041	33.5695	16.5950	42.9751	0.3090	0.0165
34.3158	22.1750	25.2042	34.5240	16.3148	41.2062	0.3078	0.0154
34.3456	20.6359	25.0226	33.7643	16.7667	40.7152	0.3158	0.0159

حال برای کلاس صفر داریم:

$$Height0: \mu = 34.30, \sigma^2 = 20.91 \rightarrow f_X(x) = \frac{1}{\sqrt{41.8\pi}} \exp\left(-\frac{(x-34.30)^2}{67.14}\right)$$

$$Weight0: \mu = 24.90, \sigma^2 = 33.57 \rightarrow f_X(x) = \frac{1}{\sqrt{67.14\pi}} \exp\left(-\frac{(x-24.90)^2}{67.14}\right)$$

$$\text{Bark Days: } np = \text{mean} \rightarrow 30p = 16.595 \rightarrow p = 0.553 \rightarrow P(X = k) = \binom{30}{k} (0.553)^k (0.447)^{30-k}$$

$$\text{Ear head ratio: } \mu = 0.309 \text{ var} = 0.0165 \rightarrow \frac{(a+b)}{2} = 0.3090, \frac{(b-a)^2}{12} = 0.0165$$

$$\rightarrow a = 0.087, b \cong 0.532 \rightarrow f_X(x) = \frac{1}{b-a} = \frac{1}{0.532-0.087} = \frac{1}{0.445} \cong 2.247; a \leq x \leq b$$

برای کلاس 1 داریم:

$$\text{Height: } \mu = 34.3158, \text{ var} = 22.175 \rightarrow f_X(x) = \frac{1}{\sqrt{44.35\pi}} \exp\left(-\frac{(x-34.316)^2}{44.35}\right)$$

$$\text{Weight: } \mu = 25.2043, \text{ var} = 34.520 \rightarrow f_X(x) = \frac{1}{\sqrt{69.04\pi}} \exp\left(-\frac{(x-25.204)^2}{69.04}\right)$$

$$\text{Bark days: } np = \text{mean} \rightarrow 30p = 16.314 \rightarrow p = 0.543 \rightarrow P(X = k) = \binom{30}{k} (0.543)^k (0.457)^{30-k}$$

$$\text{Ear head ratio: } \mu = 0.3078 = \frac{a+b}{2}, \text{ var} = 0.0154 = \frac{(b-a)^2}{12} \rightarrow a = 0.097, b = 0.520 \rightarrow f_X(x) = \frac{1}{0.423} \cong 2.364; a \leq x \leq b$$

برای کلاس 2 نیز چنین است:

$$\text{Height: } \mu = 34.346, \text{ var} = 20.636 \rightarrow f_X(x) = \frac{1}{\sqrt{41.272\pi}} \exp\left(-\frac{(x-34.346)^2}{41.272}\right)$$

$$\text{Weight: } \mu = 25.023, \text{ var} = 33.764 \rightarrow f_X(x) = \frac{1}{\sqrt{67.53\pi}} \exp\left(-\frac{(x-25.023)^2}{67.53}\right)$$

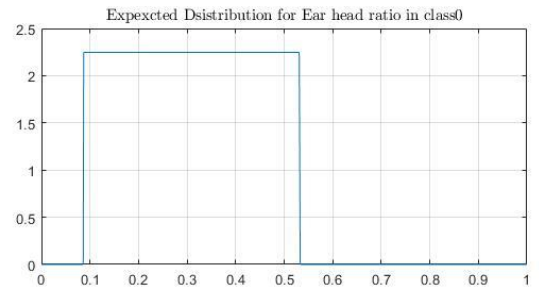
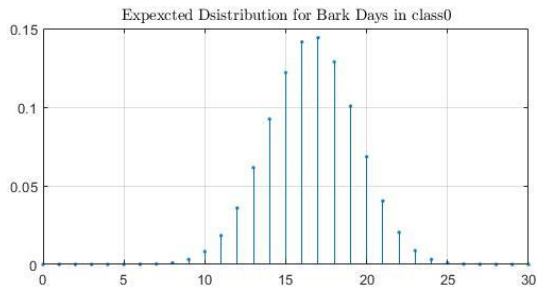
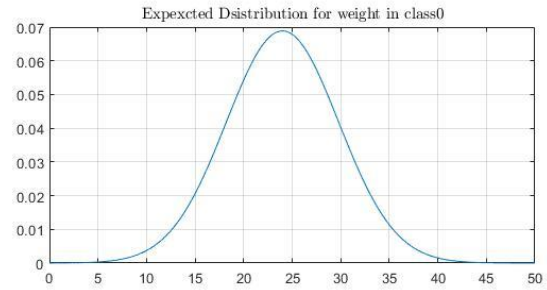
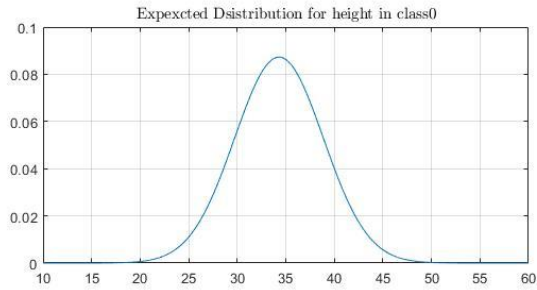
$$\text{Bark days } np = \text{mean} \rightarrow 30p = 16.766 \rightarrow p = 0.559 \rightarrow P(X = k) = \binom{30}{k} (0.559)^k (0.441)^{30-k}$$

$$\text{Ear head ratio: } \mu = 0.3158 = \frac{a+b}{2}, \text{ var} = 0.0159 = \frac{(b-a)^2}{12} \rightarrow b = 0.535, a = 0.098 \rightarrow f_X(x) = \frac{1}{b-a} = \frac{1}{0.535-0.098} \cong 2.288; a \leq x \leq b$$

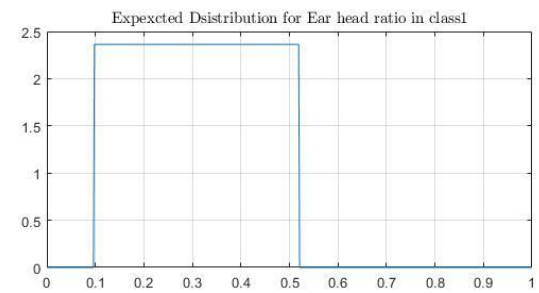
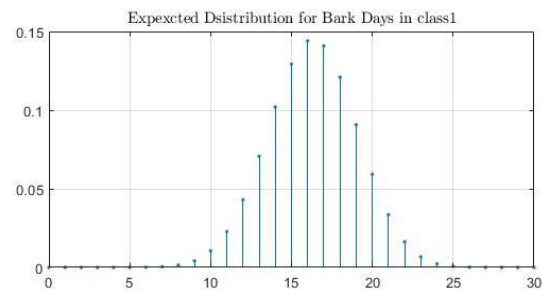
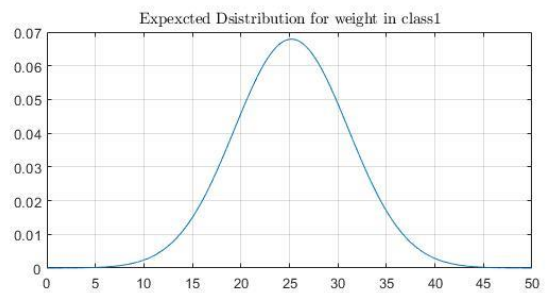
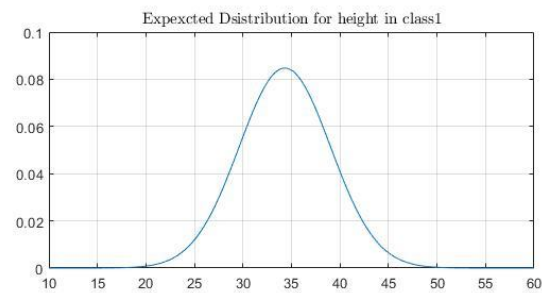
حال توزیع های بدست آمده را به کد اصلی اضافه می کنیم و تمامی آنها را رسم می کنیم.

توزیع های بدست آمده توسط روابط فوق:

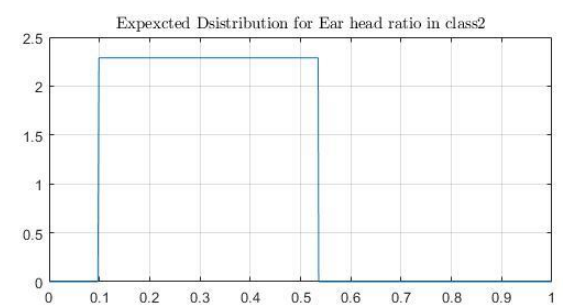
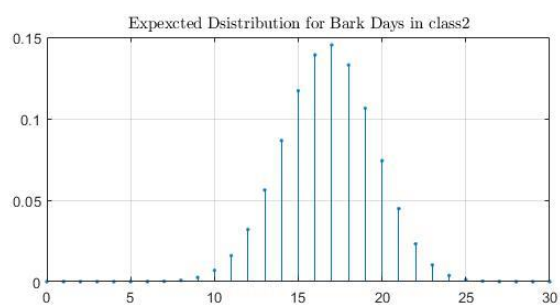
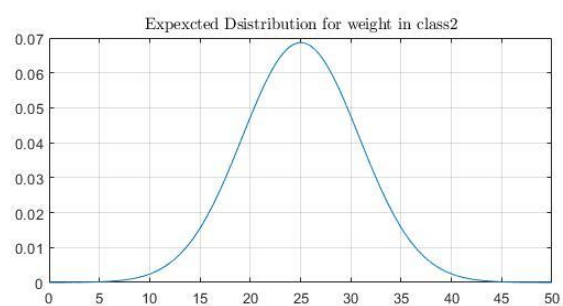
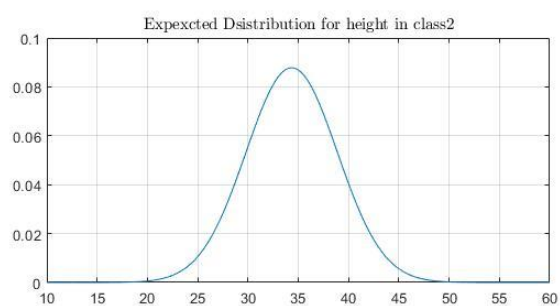
برای کلاس 0 داریم:



برای کلاس 1 داریم:



برای کلاس 2 داریم:



کد 1 simulation

```

clc
m=[height1 weight bark_days ear_head_ratio class1];
N=zeros(1,3);
for i=1:length(height1)-1
    for j=1:length(height1)-1
        if m(j,5)>m(j+1,5)
            a=m(j,5);
            b=m(j+1,5);
            m(j,5)=b;
            m(j+1,5)=a;
        end
    end
end
for i=1:length(height1)
    if m(i,5)==0
        N(1,1)=N(1,1)+1;
    elseif m(i,5)==1
        N(1,2)=N(1,2)+1;
    else
        N(1,3)=N(1,3)+1;
    end
end
%%
h0=[]; h1=[]; h2=[];
w0=[]; w1=[]; w2=[];
b0=[]; b1=[]; b2=[];
e0=[]; e1=[]; e2=[];
c0=[]; c1=[]; c2=[];

for i=1:length(height1)
    if i<=N(1,1)
        h0=[h0 m(i,1)]; w0=[w0 m(i,2)]; b0=[b0 m(i,3)]; e0=[e0 m(i,4)];
        c0=[c0 m(i,5)];
    elseif i>N(1,1)&i<=(N(1,2)+N(1,1))
        h1=[h1 m(i,1)]; w1=[w1 m(i,2)]; b1=[b1 m(i,3)]; e1=[e1 m(i,4)];
        c1=[c1 m(i,5)];
    else
        h2=[h2 m(i,1)]; w2=[w2 m(i,2)]; b2=[b2 m(i,3)]; e2=[e2 m(i,4)];
        c2=[c2 m(i,5)];
    end
end
PM=zeros(3,8);
PM(1,1)=mean(h0); PM(1,2)=var(h0); PM(1,3)=mean(w0); PM(1,4)=var(w0);
PM(1,5)=mean(b0);
PM(1,6)=var(b0); PM(1,7)=mean(e0); PM(1,8)=var(e0);
PM(2,1)=mean(h1); PM(2,2)=var(h1); PM(2,3)=mean(w1); PM(2,4)=var(w1);
PM(2,5)=mean(b1);
PM(2,6)=var(b1); PM(2,7)=mean(e1); PM(2,8)=var(e1);
PM(3,1)=mean(h2); PM(3,2)=var(h2); PM(3,3)=mean(w2); PM(3,4)=var(w2);
PM(3,5)=mean(b2);
PM(3,6)=var(b2); PM(3,7)=mean(e2); PM(3,8)=var(e2);
%%
%class 0
n0=min(b0):1:max(b0);

```



```

N0=zeros(1,length(n0));
for i=1:length(b0)
    for j=1:length(n0)
        if b0(i)==n0(j);
            N0(j)=N0(j)+1;
        end
    end
end

end
subplot(2,2,1)
histogram(h0,20)
grid
title('Class 0 height','interpreter','latex')
subplot(2,2,2)
histogram(w0,20)
title('Class 0 weight','interpreter','latex')
grid
subplot(2,2,3)
stem(n0,N0,'.')
title('Class 0 bark days','interpreter','latex')
grid
subplot(2,2,4)
histogram(e0,20)
grid
title('Class 0 ear head ratio','interpreter','latex')
%%
n1=min(b1):1:max(b1)
N1=zeros(1,length(n1))
for i=1:length(b1)
    for j=1:length(n1)
        if b1(i)==n1(j);
            N1(j)=N1(j)+1;
        end
    end
end

end

figure
subplot(2,2,1)
histogram(h1,20)
grid
title('Class 1 height','interpreter','latex')
subplot(2,2,2)
histogram(w1,20)
title('Class 1 weight','interpreter','latex')
grid
subplot(2,2,3)
stem(n1,N1,'.')
title('Class 1 bark days','interpreter','latex')
grid
subplot(2,2,4)
histogram(e1,20)
grid
title('Class 1 ear head ratio','interpreter','latex')
%%
figure

```

```

n2=min(b2):1:max(b2);
N2=zeros(1,length(n2));
for i=1:length(b2)
    for j=1:length(n2)
        if b2(i)==n2(j);
            N2(j)=N2(j)+1;
        end
    end
end

end
subplot(2,2,1)
histogram(h2,20)
grid
title('Class 2 height','interpreter','latex')
subplot(2,2,2)
histogram(w2,20)
title('Class 2 weight','interpreter','latex')
grid
subplot(2,2,3)
stem(n2,N2,'.')
title('Class 2 bark days','interpreter','latex')
grid
subplot(2,2,4)
histogram(e2,20)
grid
title('Class 2 ear head ratio','interpreter','latex')
%%
fprintf('\n\nThe output mean and variance matrix is as follow:\n')
PM
%%
%The expected distributions for class 0
figure
%height
subplot(2,2,1)
t=10:1e-2:60;
fxh0=1/sqrt(41.8*pi)*exp(-(t-34.30).^2/41.8);
plot(t,fxh0)
grid
title('Expexcted Dsistribution for height in class0','interpreter','latex')
%weight
subplot(2,2,2)
t=0:1e-2:50;
fxw0=1/sqrt(67.14*pi)*exp(-(t-24.00).^2/67.14);
plot(t,fxw0)
grid
title('Expexcted Dsistribution for weight in class0','interpreter','latex')
%barkly days
subplot(2,2,3)
n0=30;
p0=0.553;
x0=0:30;
P0=zeros(1,length(x0));
for i=1:length(x0)
    P0(i)=nchoosek(n0,x0(i)).*(p0.^x0(i)).*((1-p0).^(30-x0(i)));
end
stem(x0,P0,'.')

```

```

title('Expexcted Dsistribution for Bark Days in
class0','interpreter','latex')
grid
%ear head ratio
subplot(2,2,4)
t=0:1e-3:1;
a0=0.087;
b0=0.532;
fxe0=1/(b0-a0)*(heaviside(t-a0)-heaviside(t-b0));
plot(t,fxe0)
grid
title('Expexcted Dsistribution for Ear head ratio in
class0','interpreter','latex')

%%
%The expected distributions for class 1
figure
%height
subplot(2,2,1)
t=10:1e-2:60;
fxh1=1/sqrt(44.35*pi)*exp(-(t-34.316).^2/44.35);
plot(t,fxh1)
grid
title('Expexcted Dsistribution for height in class1','interpreter','latex')
%weight
subplot(2,2,2)
t=0:1e-2:50;
fxw1=1/sqrt(69.04*pi)*exp(-(t-25.204).^2/69.04);
plot(t,fxw1)
grid
title('Expexcted Dsistribution for weight in class1','interpreter','latex')
%barkly days
subplot(2,2,3)
n1=30;
p1=0.543;
x1=0:30;
P1=zeros(1,length(x1));
for i=1:length(x1)
    P1(i)=nchoosek(n1,x1(i)).*(p1.^x1(i)).*((1-p1).^(30-x1(i)));
end
stem(x1,P1, '.')
title('Expexcted Dsistribution for Bark Days in
class1','interpreter','latex')
grid
%ear head ratio
subplot(2,2,4)
t=0:1e-3:1;
a1=0.097;
b1=0.520;
fxe1=1/(b1-a1)*(heaviside(t-a1)-heaviside(t-b1));
plot(t,fxe1)
grid
title('Expexcted Dsistribution for Ear head ratio in
class1','interpreter','latex')
%%
%The expected distributions for class 2
figure

```

```

%height
subplot(2,2,1)
t=10:1e-2:60;
fxh2=1/sqrt(41.272*pi)*exp(-(t-34.346).^2/41.272);
plot(t,fxh2)
grid
title('Expexcted Dsistribution for height in class2','interpreter','latex')
%weight
subplot(2,2,2)
t=0:1e-2:50;
fxw2=1/sqrt(67.53*pi)*exp(-(t-25.023).^2/67.53);
plot(t,fxw2)
grid
title('Expexcted Dsistribution for weight in class2','interpreter','latex')
%barkly days
subplot(2,2,3)
n2=30;
p2=0.559;
x2=0:30;
P2=zeros(1,length(x2));
for i=1:length(x2)
    P2(i)=nchoosek(n2,x2(i)).*(p2.^x2(i)).*((1-p2).^(30-x2(i)));
end
stem(x2,P2,'.')
title('Expexcted Dsistribution for Bark Days in
class2','interpreter','latex')
grid
%ear head ratio
subplot(2,2,4)
t=0:1e-3:1;
a1=0.098;
b1=0.535;
fxe2=1/(b1-a1)*(heaviside(t-a1)-heaviside(t-b1));
plot(t,fxe2)
grid
title('Expexcted Dsistribution for Ear head ratio in
class2','interpreter','latex')
%%

```

alg 1 کد

```

clc
a=zeros(1,4);
t=1;
while t==1
    a(1,1)=input('\nEnter the height\t');
    a(1,2)=input('\nEnter the weigth\t');
    a(1,3)=input('\nEnter the bark days\t');
    a(1,4)=input('\nEnter the ear head ration\t');
    %%%class 0%%
    h0=1/sqrt(41.8*pi)*exp(-(a(1,1)-34.30).^2/41.8);
    w0=1/sqrt(67.14*pi)*exp(-(a(1,2)-24.00).^2/67.14);
    a(1,3)=round(a(1,3));
    n0=30;

```

```

p0=0.553;
b0=nchoosek(n0,a(1,3)).*(p0.^a(1,3)).*((1-p0).^(30-a(1,3)));
aa0=0.087;
bb0=0.532;
e0=1/(bb0-aa0)*(heaviside(a(1,4)-aa0)-heaviside(a(1,4)-bb0));
m0=h0*w0*b0*e0*(1200/3450);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%class 1%%
h1=1/sqrt(44.35*pi)*exp(-(a(1,1)-34.316).^2/44.35);
w1=1/sqrt(69.04*pi)*exp(-(a(1,2)-25.204).^2/69.04);
n1=30;
p1=0.543;
b1=nchoosek(n1,a(1,3)).*(p1.^a(1,3)).*((1-p1).^(30-a(1,3)));
aa1=0.097;
bb1=0.520;
e1=1/(bb1-aa1)*(heaviside(a(1,4)-aa1)-heaviside(a(1,4)-bb1));
m1=h1*w1*b1*e1*(1350/3450);
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%%class 2%%
h2=1/sqrt(41.272*pi)*exp(-(a(1,1)-34.346).^2/41.272);
w2=1/sqrt(67.53*pi)*exp(-(a(1,2)-25.023).^2/67.53);
n2=30;
p2=0.559;
b2=nchoosek(n2,a(1,3)).*(p2.^a(1,3)).*((1-p2).^(30-a(1,3)));
aa2=0.098;
bb2=0.535;
e2=1/(bb2-aa2)*(heaviside(a(1,4)-aa2)-heaviside(a(1,4)-bb2));
m2=h2*w2*b2*e2*(900/3450);
m=max(max(m0,m1),max(m1,m2));
if m==m0
    fprintf('\nThe input data is assigned to %d',0)
elseif m==1
    fprintf('\nThe input data is assigned to %d',1)
else
    fprintf('\nThe input data is assigned to %d',2)
end
[m0 m1 m2]
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

t=input('\nDo you want to continue?(1/0)\t');

end

```

نمونه ورودی خروجی:

Enter the heightt36.710

Enter the weightt21.140

Enter the bark dayst26

Enter the ear head ratio t 0.163527

The input data is assigned to 0

ans =

1.0×10^{-6} *

0.8099 0.5601 0.7059

Do you want to continue? (1/0) t