**پروژه هفتم:**

## بازشناسی آماری الگو

در این تمرین سعی شده است با توجه به dataset استخراج شده از uci اقدام به تصمیم گیری بیزی و کاهش بعد داده شود . این ردیف داده مربوط است به اندازه گیری نرخ ضربان جنین و مشخه های انقباض رحمی روی کاردیوگرام که توسط متخصص جمع آوری شده است . در انجام این تمرین تعداد ۳۰۰ نمونه از ۲۱۲۶ انتخاب شود.پس از استخراج و انتخاب ویژگی از ۲۱ مشخصه اقدامات زیر برروی آنها انجام می شود:
classification naïve baysian ,Non parametric estimation (parzen &kNN),Feature extraction(PCA),Feature selection (randomized selection)
که در محیط MATLAB با استفاده از PRTOOLS انجام می شود.و در نهایت با توجه به نتایج بدست آمده نتیجه گیری انجام می شود.

**معرفی داده:**

در این تمرین مجموعه داده ها از uci استخراج شده است. که مربوط است به اندازه گیری نرخ ضربان جنین و مشخصه های انقباض رحمی روی کاردیوگرام که توسط متخصص جمع آوری شده است.
در این مجموعه داده 21 مشخصه وجود دارد که با توجه به نوع تمرین که استخراج و انتخاب مشخصه در آن وجود دارد این مقدار مناسب و کافی به نظر می آید از طرفی با توجه به تعداد بالای نمونه ( 2126 ) تعداد 300 نمونه در انجام این گزارش کافی به نظر می رسد.
(3خصوصیات آماری داده ها:
در این گزارش با استفاده ازMatlab Toolbox به محاسبه مشخصات آماری داده ها پرداخته شده است . مقادیر آن در داده ها گزارش جدول زیر آمده است:

Min= فرکانس پایین هیستوگرام
Max= فرکانس بالای هیستوگرام
NMax= تعداد پیک های هیستوگرام
Nzero= تعداد صفرهای هیستوگرام
Mode= مد هیستوگرام
Mean= متوسط هیستوگرام
Median = میانه هیستوگرام
Variance= واریانس هیستوگرام

www.matlabproject.ir

| Features | chi-square |
|---|---|
| LB | 243.5 |
| AC | 1701.2 |
| FM | 84.3 |
| UC | 253.8 |
| DL | 1336.1 |
| DS | 96.4 |
| DP | 975 |
| ASTV | 450.9 |
| MSTV | 872.1 |
| ALTV | 879.8 |
| MLTV | 240.9 |
| Width | 842.4 |
| Min | 897 |
| Max | 471.4 |
| Nmax | 509.4 |
| Nzeros | 212.8 |
| Mode | 349 |
| Mean | 524.4 |
| Median | 416.9 |
| Variance | 1279.2 |
| Tendency | 143.9 |

www.ma

## روش کار:

### نظریه تصمیم گیری بیضی:

تئوری بیضی یک روش آماری اساسی برای حل مشکلات شناسایی الگو است.این روش بین تصمیم کلاس‌های مختلف و هزینه‌ای که این تصمیم‌ها در بر دارند مصالحه‌ای برقرار می‌کند و سپس بهترین را برمی‌گزینند.بنابراین برای این نحوه‌ی تصمیم گیری نیاز است که توابع توزیع احتمالی و مقادیر متناسب با آنها مشخص باشند.این نظریه به صورت ریاضی توسط رابطه زیر بیان می‌شود:

$$P(w_i \mid x) = P(x \mid w_i) . P(w_i) / P(x)$$

$$P(x)=\sum_{j=1}^{c} P(x|W_j)P(w_j)$$

$P(x \mid w_i)$، Conditional Density، تابع توزیع احتمال برای ویژگی X ،زمانی که در کلاس $w_i$ قرار دارد و $P\ (w_i)$، Prior Probability را بیان می‌کند.(C تعداد کلاس‌ها را نشان می‌دهد).

روش‌های پارامتری به کار رفته در گزارش

اگر $P(x \mid w_i)$ و $P\ (w_i)$ مشخص باشند با کمک تئوری بیض و با توجه به $P(w_i \mid x)$ بزرگتر می‌توان تصمیم گیری کرد و یک طبقه بندی کننده بهینه تهیه کرد ولی متاسفانه در بسیاری از کاربردهای شناسایی آماری الگو ، بندرت اطلاعات کاملی از ساختار احتمالی مسئله پیش رو است و از این رو نیاز به روش‌های مختلف برای تخمین این توابع توزیع روشن می‌گردد.با کمک داده‌هایی که به عنوان داده‌های آموزش تهیه شده‌اند می توان این مشکل را حل کرد.درکاربردهای شناسایی الگو همراه با سرپرستی اگر طول داده‌ها خیلی کم نباشند و ابعاد بردار ویژگی، زیاد نباشد می‌توان تابع $P(x \mid w_i)$ *را با دو روش پارامتری* Maximum Liklihood و Baysian Estimation *را تخمین زد.*

در هر دو روش *ML* و تخمین بیضی توزیع $P(x \mid w_i)$ معلوم فرض می‌گردد ولی پارامترهای آن نامشخص هستند، از این رو باید این *پارامترها* تخمین زده شوند. در روش MLفرض می‌شود پارامتری که باید تخمین زده شود ثابت است و با توجه به استقلال ویژگی‌ها و مجموعه D که شامل داده‌های آموزشی می‌باشد رابطه زیر صادق است:

$$P(D|\theta) = \prod_{k=1}^{n} P(x_k|\theta) = F(\theta)$$

$$P(D|\theta)\ is\ called\ the\ likelihood\ of\ \theta$$

می‌توان با صفر قراردادن گرادیان تابع $F(\theta)$ مقدار مناسب پارامتری که این تابع را ماکزیمم می‌کند محاسبه کرد. بنابراین به کمک داده‌های آموزش بهترین $\theta$ که این تابع را ماکزیمم می‌کند تخمین زد(حتی با لگاریتم گرفتن هم می‌توان حاصلضرب را به حاصلجمع تبدیل کرد تا محاسبات ساده‌ترشود.)ولی در روش تخمین بیضی $\theta$ رندوم اختیار می‌گردد و این روش دارای پیچیدگی محاسباتی بیشتری نسبت به ML می‌باشد. بنابراین توزیع $P(x \mid w_i)$ معلوم فرض می‌گردد و اطلاعاتی که در مورد$\theta$ وجود دارد از P($\theta$) معلوم و اطلاعات موجود در داده‌های آموزش حاصل می‌گردد.

رابطه زیر از داده‌های آموزش حاصل می‌گردد:

$$P(D|\theta) = \prod_{k=1}^{n} P(x_k|\theta)$$

و سپس طبق قاعده بیض روابط زیر محاسبه می‌گردد:

$$P(D|\theta) = \frac{P(D|\theta).P(\theta)}{\int P(D|\theta).P(\theta)d\theta}$$

www.matlabproject.ir

$$P(x|D) = \int P(x|\mu).P(\mu|D)d\mu \; is \; Gaussian$$

و سپس از روی $P(x|D)$ و قاعده بیض توزیع زیر حاصل می گردد.

$$P(w_i|x,D) = \frac{P(X|w_i,D_i).P(w_i)}{\sum_{j=1}^{c} P(X|w_j,D).P(w_j)}$$

$$\max_{w_j}[P(w_j|x,D)] \equiv \max_{w_j}[P(X|w_i,D_j).P(w_j)]$$

متون آماری مربوط به تحلیل ممیزی بیضی را می توان به دو نوع تقسیم کرد. نوع اول روش شبه بیضی است که هوبرت و ون دریسن آن را به کار برده اند و نوع دوم استفاده از پیشین آگاهی نابخش ناسره است که توسط گیسرر و پرس انجام گرفته است. در روش شبه بیضی پارامترهای جوامع رقیب با روش درستنمایی ماکزیمم(ML) برآورد می شوند و سپس احتمال های پسین تعلق یک مشاهده جدید به هر یک از $k$ جامعه رقیب با فرمول بیض محاسبه می شود. روش شبه بیضی را بعضاً روش درستنمایی ماکزیمم تعمیم یافته می نامند. در روش های پیشین آگاهی نابخش، برای پارامترهای جوامع رقیب پیشین های آگاهی نابخش در نظر گرفته می شود. به استثناء پرس که اشاره ای به پیشین آگاهی بخش مزدوج نموده است متون آماری به این مقوله نپرداخته اند. از طریق این نوع پیشین است که امکان مداخله اطلاعات خارج از حیطه ی داده ها میسر می شود. به‌طور کلی تئوری بیضی یک روش آماری برای شناسایی الگوی آماری است. تکنیکهای آماری به دو دسته پارامتریک و غیرپارامتریک تقسیم‌بندی می‌شوند.

## آموزش پارامتری و غیرپارامتری:

روشهای پارامتریک:

در این روش عموماً فرض بر این قرار است که داده‌ها توزیع نرمال دارد و باید پارامترهای توزیع را تخمین زد. مراحل شبیه سازی در حالت کلی برای گروه تیروئید در روش پارامتری به صورت زیر می باشد:

ابتدا توسط دستور dataset داده ها در کلاسهای مختلف طبقه بندی می شوند سپس با استفاده از دستور gendat داده های فوق به دو گروه آموزش و تست تقسیم شده در مرحله بعد برای بررسی ویژگیها و انتخاب بهترین ویژگی با استفاده از توسط دستور plotf ، توزیع هر ویژگی رسم می شود. (محور x مقدار ویژگی و محور y $P(x|\omega_i)$ را به ازای کلاس های مختلف(i=1,2,3) نشان می دهد. در مراحل نهایی برای طبقه بندی داده ها با توجه به روش نزدیکترین همسایگی (ML) Max likelihood از دستور loglc و استفاده شده که یک طبقه بندی کننده خطی است و توسط ML معیار کارآیی ماکزیمم می شود و برای طبقه بندی داده ها با روش تخمین بیضین baysian estimation از دستور nbayesc استفاده

www.matlabproject.ir

شد که پارامترهای ورودی آن میانگین و کوواریانس داده های آموزش می باشد استفاده می شود پس از دستورهای loglc و nbayesc خطای طبقه بندی کننده محاسبه می گردد.

**روشهای غیرپارامتریک:**

در این روش هیچگونه فرضی در مورد توزیع داده‌ها نمی‌شود ولی در انتها با کمک داده‌های آموزشی می‌توان تابع توزیع را تخمین زد. روش های غیر پارامتریک مطرح شده در این گزارش شامل روش k-nearest neighbor و parzen می باشد. مراحل شبیه‌سازی در حالت کلی در روش پارامتری به صورت زیر می باشد:

در روش دسته بندیk-nearest neighbor estimation از دستور Knnc استفاده می شود. ابتدا توسط داده های train به طبقه بندی کننده آموزش داده شد و سپس با استفاده از داده های تست و دستور testk میزان خطای طبقه بندی کننده knn محاسبه گردید تا از صحت عملکرد آن در دسته بندی داده ها مطمئن شویم.

در روش دسته بندیparzen window estimation نیز، از دستورparzenc استفاده شد . ابتدا توسط داده های train به طبقه بندی کننده آموزش داده شد و سپس با استفاده از داده های تست و دستورtestp میزان خطای طبقه بندی کننده parzen محاسبه گردید تا از صحت عملکرد آن در دسته بندی داده ها مطمئن شویم .

روش parzen را می توان توسط Probability Neural Network هم پیاده سازی کرد که الگوها به همراه d ویژگی مورد نظر توسط c کلاس نمونه برداری می شود.این روش به یک مقدار ثابتspread وابسته است که هرچقدر این ثابت بزرگتر باشد مقدار این تخمین هموارتر خواهد بود.ولی از طرف دیگر به تعداد نرون های زیادی برای تشکیل تابع با سرعت تغییر بالا نیاز خواهد بود.با استفاده از مقدار های مختلف می توان حالت بهینه را برای یک مسئله خاص پیدا کرد.

برای طبقه بندی داده ها به روش pnn از یکی از toolbox های شبکه عصبی به نام "nntool" استفاده می شود.تابع newpnn با ثابت های spread متفاوت، الگوریتم طبقه بندی را اجرا می کند.

www.matlabproject.ir

## استخراج ویژگی:

مرحله استخراج ویژگی‌ها در سیستم‌های بازشناسی الگو یکی از قسمت‌های کلیدی و مؤثر در کارائی سیستم می‌باشد که دقت سیستم بازشناسی به عملکرد استخراج ویژگی وابسته است. ویژگیها باید بتوانند میان کلاسها

تمایز به وجود آورند که چون خود ویژگیها قادر به ایجاد تمایز کافی میان کلاسها نیستند، می توان از روشهای تبدیل ویژگی برای ایجاد تمایز بیشتر میان کلاسها استفاده کرد. در این بخش تعدادی از مهمترین روشهای متمایزساز خطی و غیرخطی مورد بررسی قرار می‌گیرند.

شاخه‌ای از روشهای استخراج ویژگی، روشهای خطی می‌باشند که عملکرد قابل قبولی برای دادگان با قابلیت تفکیک‌پذیری خطی دارند. روش‌های PCA و LDA از شناخته‌ترین روشهایی هستند که در این دسته قرار می‌گیرند. روش PCA روش بدون ناظر و روش LDA روش باناظر می‌باشد. هر دو روش سعی در بهینه‌سازی تابع خاصی دارند.

تبدیل خطی PCA دادگان را به فضایی نگاشت می کند که دراآن داده ها بیشترین واریانس را داشته باشند. فرض کنید دادگان شامل N مشاهده، $X_k \in R^M$ ،k=1,…,N و $\sum_{k=1}^{N} X_k = 0$ باشد. ماتریس کوواریانس مجموعه دادگان از رابطه زیر بدست می‌آید:

$$C = \frac{1}{N}\sum_{k=1}^{N} X_k X_k^T$$

با قطری‌سازی C مولفه‌های اصلی بدست می‌آیند که این مولفه‌ها تصویر متعامد روی بردارهای ویژه هستند که با حل معادله مقادیر ویژه محاسبه می‌شوند.

$$\lambda V = CV$$

که $\lambda \geq 0$ مقادیر ویژه و $V \in R^M \backslash \{0\}$ (به معنی $R^M$ به استثنای $\{0\}$) بردارهای ویژه می‌باشند. از طرفی $CV = \frac{1}{N}\sum_{k=1}^{N}(X_k . V)X_k$، همه پاسخها برای V بایستی ترکیب خطی از مشاهدات باشند، داریم:

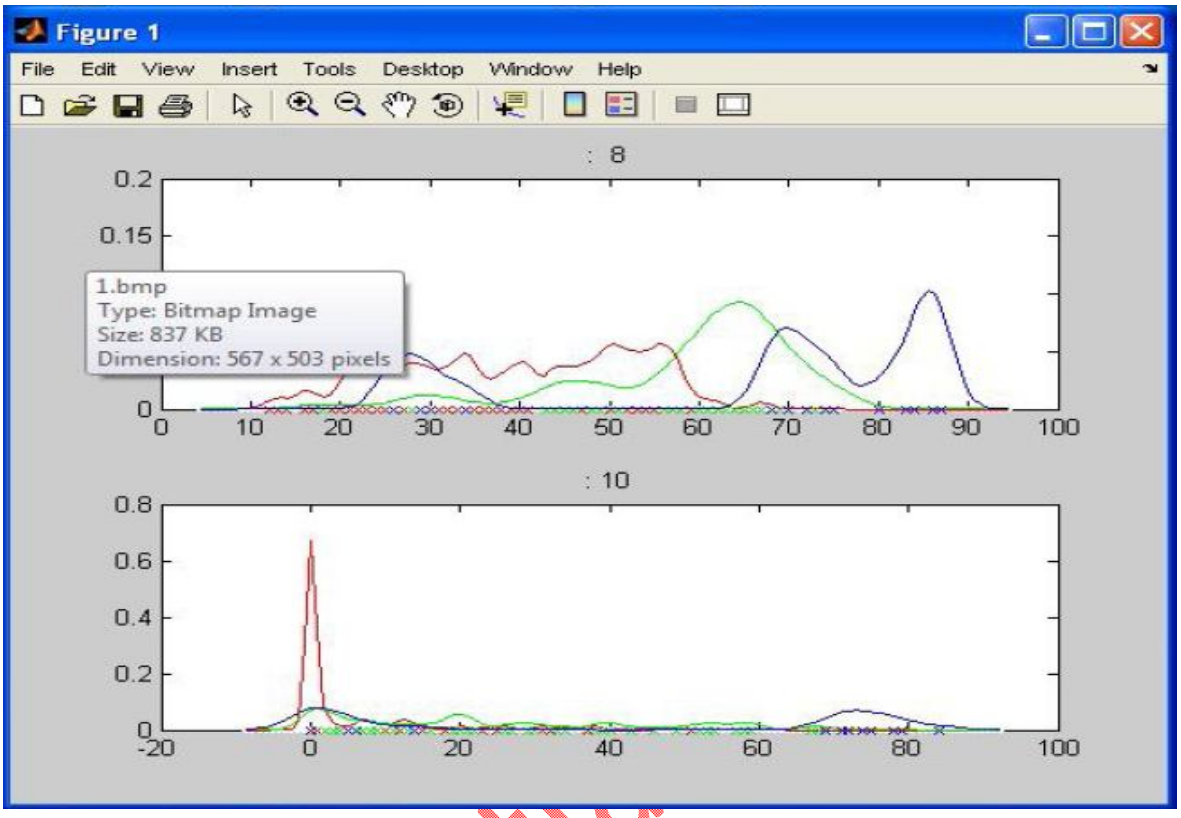
$$V = \sum_{k=1}^{N} \alpha_k X_k$$

که $\alpha_k$ها ضرایبی هستند که مقادیر آنها به گونه‌ای انتخاب می‌شوند که رابطه بالا برقرار باشد.

## شبیه‌سازی و ارائه نتایج تجربی:

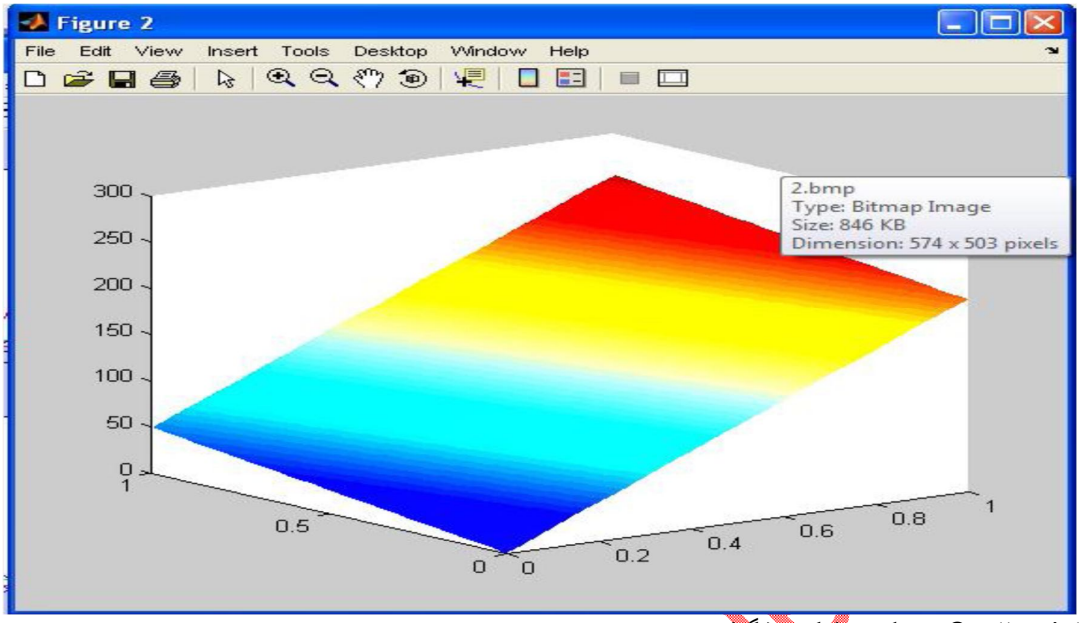
Dataset در سه کلاس با انتخاب ۲ مشخصه:

www.matlabproject.ir

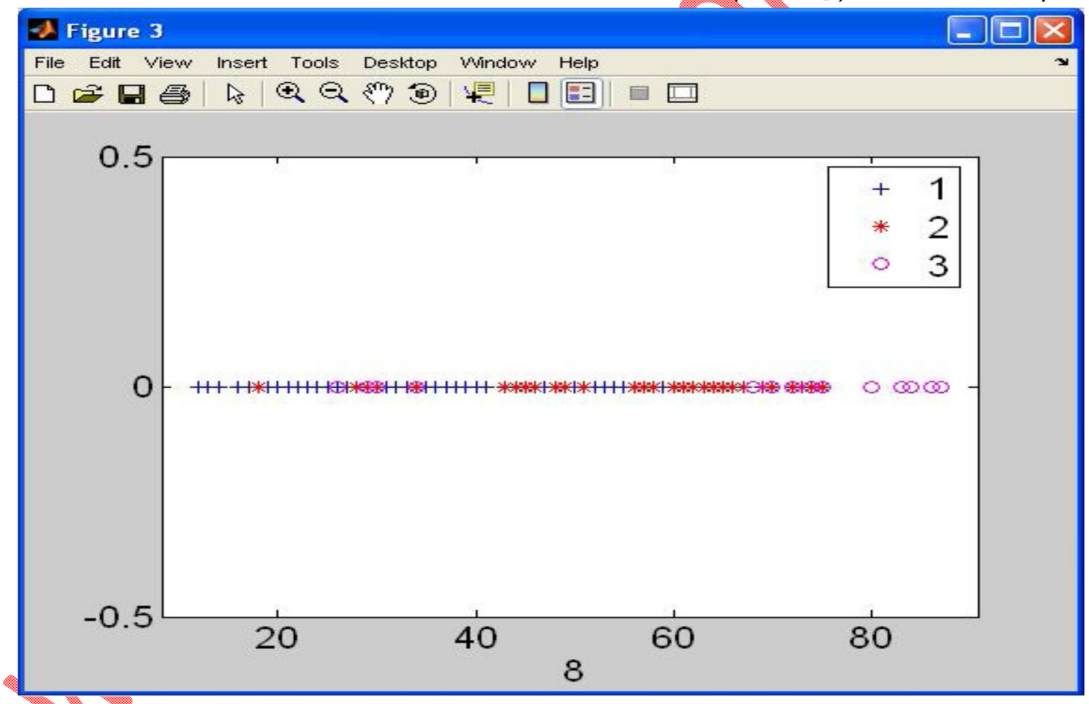



www.matla

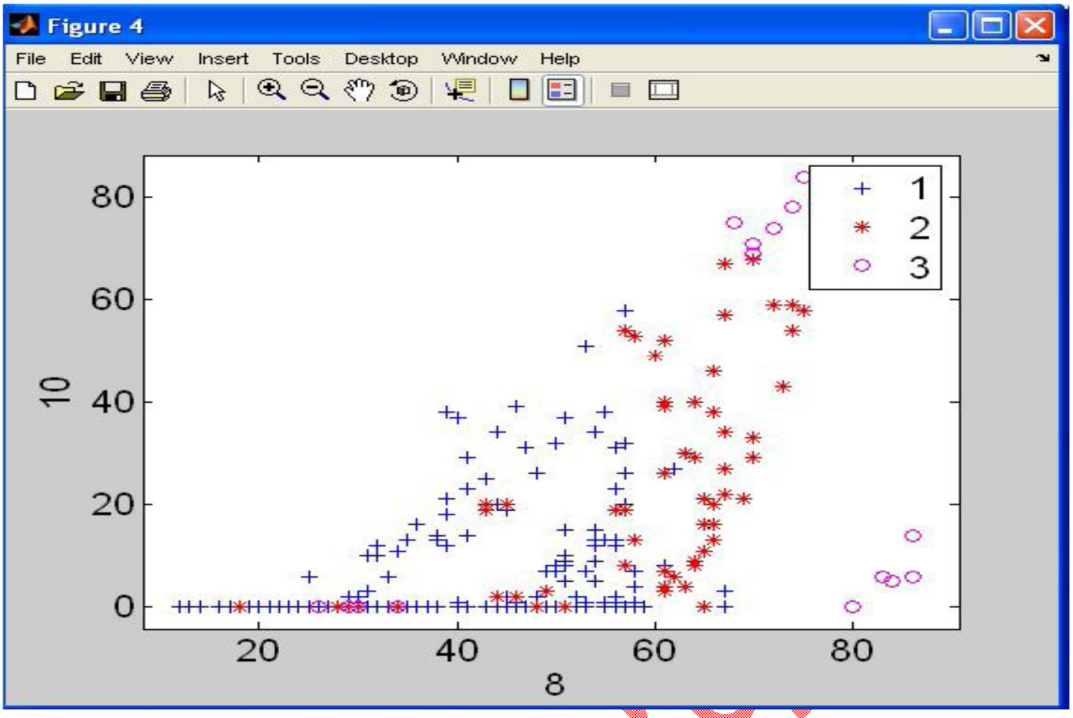
:Plot mapping in scatterplot



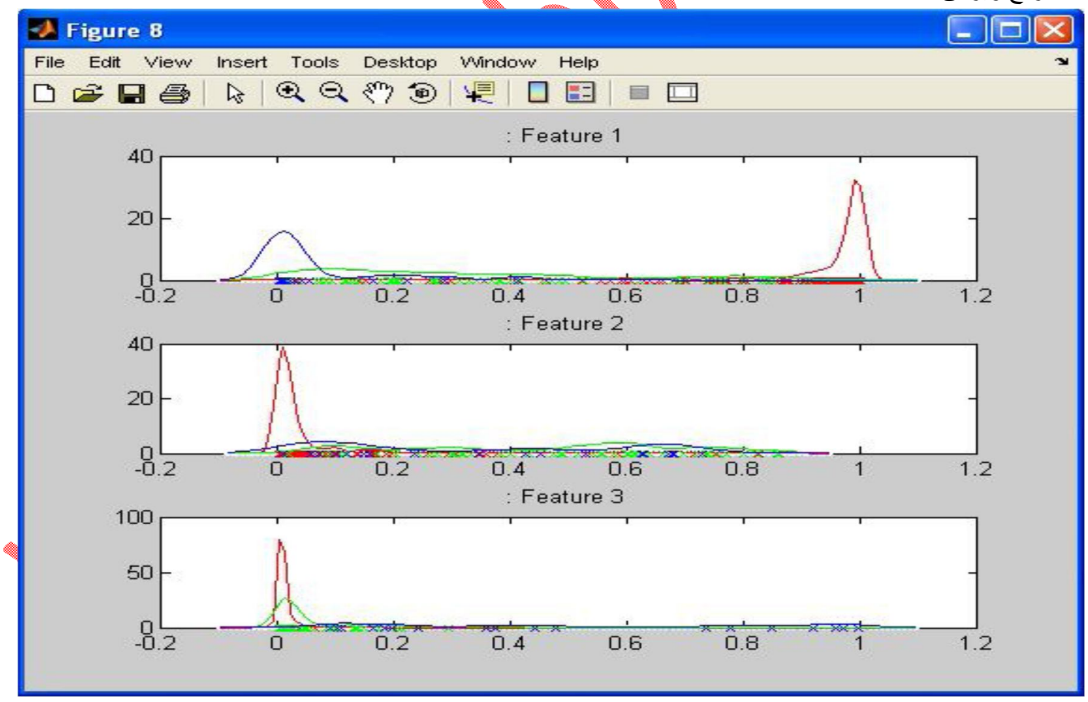

*Scatterplot* در یک بعد(یک ویژگی):



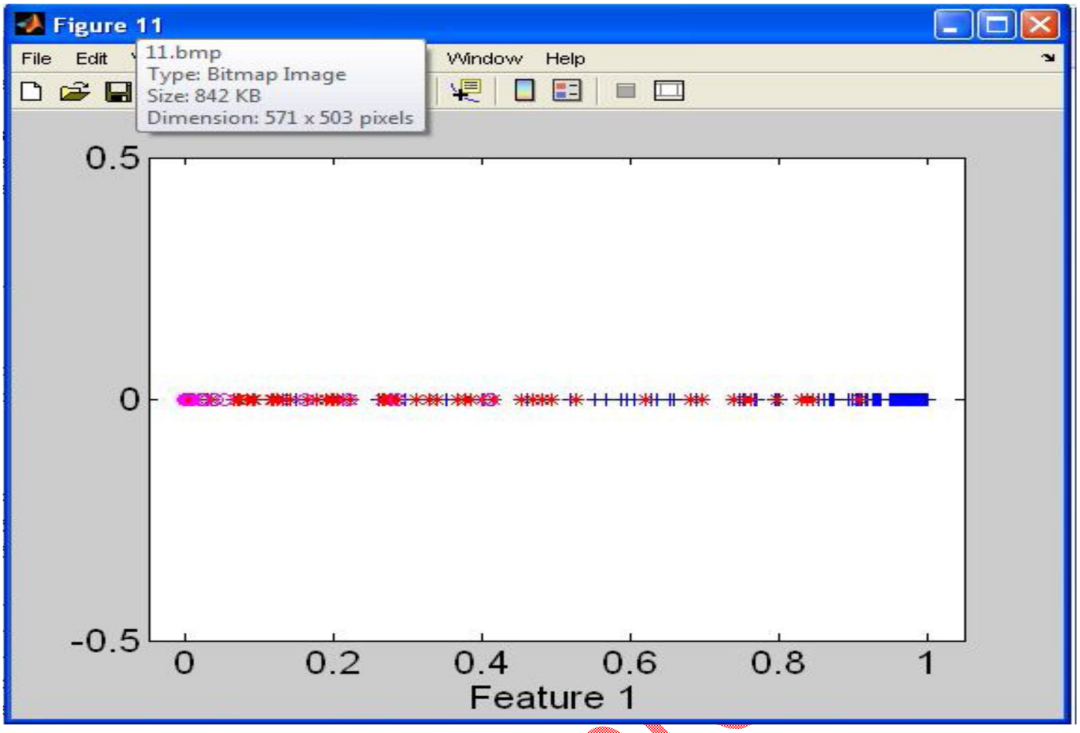

*Scatterplot* در دو بعد(دو ویژگی):



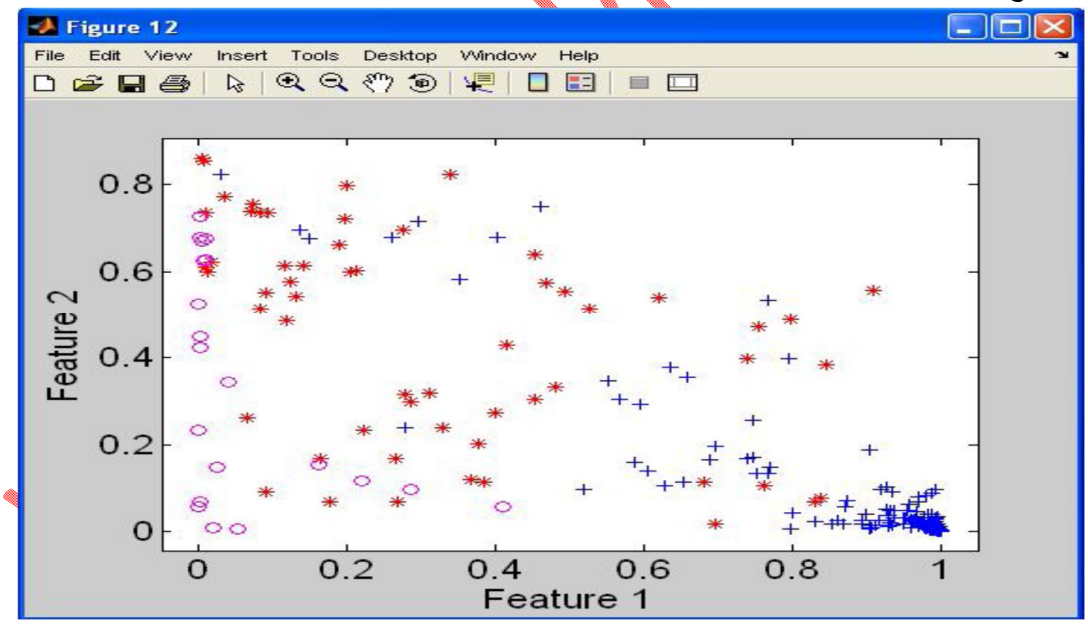

استخراج ویژگی:




استخراج ویژگی در یک بعد:




استخراج ویژگی در دو بعد:




Bayes-stimation window classifier in scatter plot در دو بعد:

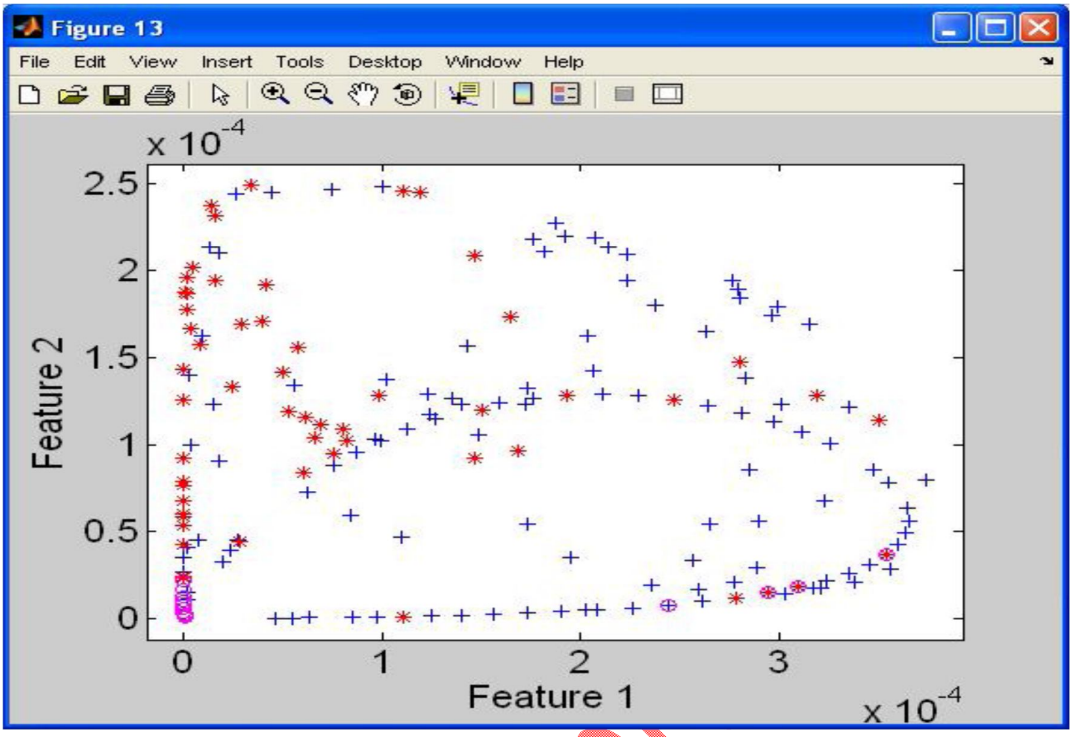
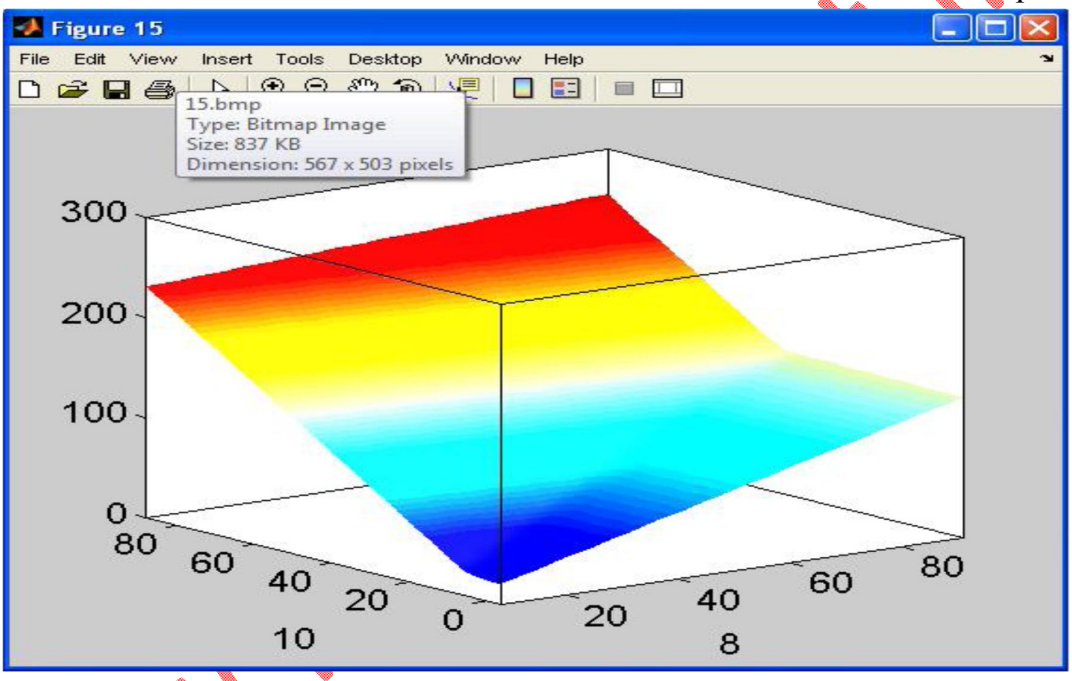


Figure 13
File Edit View Insert Tools Desktop Window Help
x 10^-4
2.5
2
1.5
1
0.5
0
Feature 2
0
1
2
3
Feature 1
x 10^-4

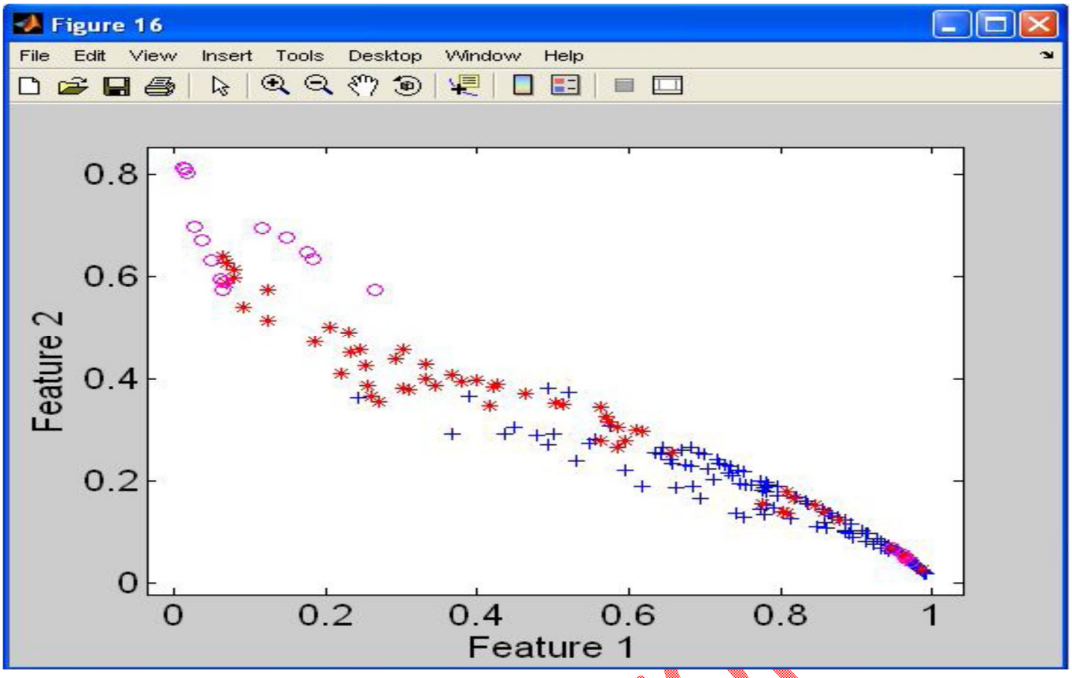
www.matlabp

:Maximum likelihood window classifier scatter 3D surface plot



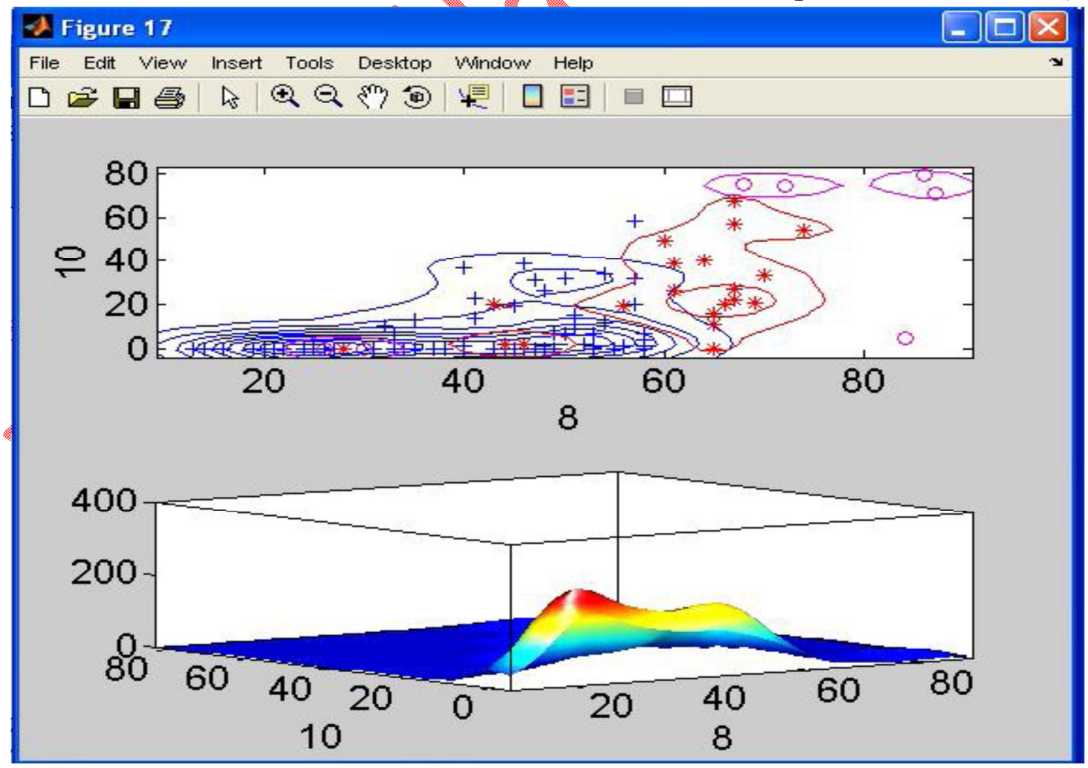

در دوبعد Maximum likelihood window classifier scatter plot:

www.



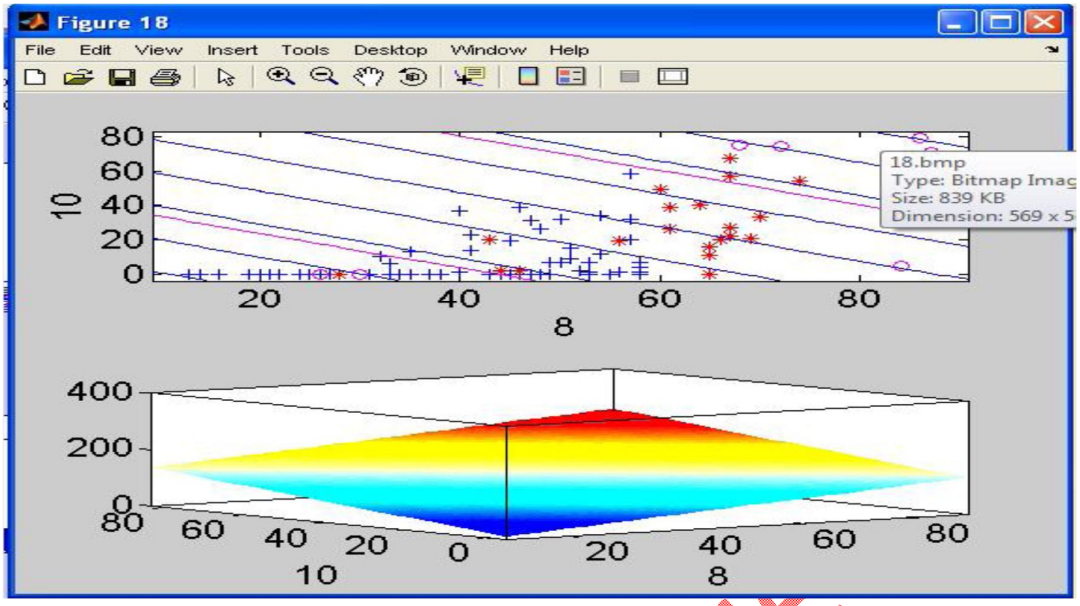

1) Parzen window classifier scatter plot

2) Parzen window classifier scatter plot in 3D surface در دوبعد




1) KNN window classifier scatter plot

2) KNN window classifier scatter plot in 3D surface در دو بعد




**کد برنامه:**

www.matlabproje

```
clc; close all; clear all;
count_b = 1;
max_b = 300;
c_1=0;
c_2=0;
c_3=0;
c_4=0;
load foetal
for i=1:501
if count_b <= max_b
if foetal(i,22) == 1 || foetal(i,22) == 2 || foetal(i,22) == 3 ||
foetal(i,22) == 4
if foetal(i,22) == 1
c_1 = c_1 + 1 ;
end
if foetal(i,22) == 2
c_2 = c_2 + 1 ;
end
if foetal(i,22) == 3
c_3 = c_3 + 1 ;
end
if foetal(i,22) == 4
c_4 = c_4 + 1 ;
end
for k=1:22
b(count_b,k) = foetal(i,k);
end
count_b = count_b+1;
end
end
```

```
end
c_1
c_2
c_3
c_4
classes = b(:,22);
features = b(:,1:20);
DS = dataset(features,classes);
fs = FEATSELF(DS,'NN',2);
fss = DS *fs;
[train , test] = gendat(fss , 0.33);
[me , co]=meancov(fss);
figure(1);plotf(fss);
figure(2);plotm(fss,3);
figure(3);scatterd(fss,1,'legend');
figure(4);scatterd(fss,2,'legend');
getfeatlab(fss)
%[W,ALF] = PCA(DS,2);
%[W,N] = PCA(DS,ALF);
%ww= DS * W;
%figure(5);plotf(ww);
%figure(6);scatterd(ww,1);
%figure(7);scatterd(ww,2)
fish = FISHERC(DS)
fishP = DS * fish;
figure(8);plotf(fishP);
figure(11);scatterd(fishP,1);
figure(12);scatterd(fishP,2);
getfeatlab(fishP)
%-------------------------------%
%bayesian method ------ classifier
%-------------------------------%
bayes_estimation = nbayesc(me , co);
temp1 = fss *bayes_estimation;
figure(13);scatterd(temp1);
%error1 = test * testc
%-------------------------------%
%maximum liklihood ------ classifier
%-------------------------------%
max_liklihood = loglc(fss);
temp2 = fss * max_liklihood;
figure(15);scatterd(fss);plotm(temp2,3);
figure(16);scatterd(temp2);
%error2 = test * testc
%-------------------------------%
%parzen window ------ classifier
%-------------------------------%
w11=parzenc(train,5);
error1 = testp(test,1)
figure(17);
subplot(2,1,1);
scatterd(train);
plotm(w11);
subplot(2,1,2);
scatterd(train);
plotm(w11,3);
%-------------------------------%
%Knn window ------ classifier
%-------------------------------%
k_n3=knnc(train,3);
k_nn3= train *k_n3;
```

www.matlabproject.ir

```
figure(18);
subplot(2,1,1);
scatterd(train);
plotm(k_nn3);
subplot(2,1,2);
scatterd(train);
plotm(k_nn3,3);
```

## جمع بندی:

در ابتدا از نظر میزان خطا، طبق گزارشات انجام شده کاملاً مشخص است که الگوی تئوری بیضی نسبت به سایر الگوها و مخصوصاً الگوهای غیرپارامتریک دارای خطای کمتریست. اما در خیلی جاها به دلیل کم بودن تعداد داده‌ها نتایج مطلوب به دست نمی‌آید. در این گزارش داده تیروئیدی از لحاظ تکنیکهای آماری پارامتریک و غیرپارامتریک بررسی شد و خطای تخمین پارامتریک و مخصوصاً خطای تخمین بیضی از سایر الگوها کمتر است که توجیه این مطلب را می‌توان به کم بودن پیچیدگی ویژگیهای آن ربط داد اما هر چه پیچیدگی ویژگیها بیشتر باشد باید به سمت استفاده از روشهای غیرپارامتریک برویم تا خطای تخمین کمتری داشته باشیم.

علاوه بر این یک مزیت طبقه بندی بیزی در مقایسه با طبقه بندی کلاسیک این است که در حالت بیضی احتمال تعلق یک مشاهده جدید به هر یک از k جامعه رقیب مشخص می شود. این در حالی است که در طبقه بندی کلاسیک آن مشاهده صرفاً طبقه بندی می شود و معلوم نیست که با چه میزانی از قاطعیت آن مشاهده طبقه بندی شده است. مزیت دیگر این است که اگر اطلاعاتی راجع به پارامترهای جوامع رقیب در خارج از محیط داده ها موجود باشد، در حالت بیزی امکان دخالت دادن آن اطلاعات وجود دارد. متون آماری مربوط به تحلیل ممیزی بیزی را می توان به دو نوع تقسیم کرد. نوع اول روش شبه بیزی است که هوبرت و ون دریسن آن را به کار برده اند و نوع دوم استفاده از پیشین آگاهی نابخش ناسره است که توسط گیسرر و پرس انجام گرفته است. در روش شبه بیضی پارامترهای جوامع رقیب با روش درستنمایی ماکزیمم برآورد می شوند و سپس احتمال های پسین تعلق یک مشاهده جدید به هر یک از k جامعه رقیب با فرمول بیز محاسبه می شود. روش شبه بیضی را بعضاً روش درستنمایی ماکزیمم تعمیم یافته می نامند. در روش های پیشین آگاهی نابخش، برای پارامترهای جوامع رقیب پیشین های آگاهی نابخش در نظر گرفته می شود. به استثناء پرس که اشاره ای به پیشین آگاهی بخش مزدوج نموده است متون آماری به این مقوله نپرداخته اند. با توجه به نقش مخربی که داده های آلوده می توانند در تحلیل چند متغیره داشته باشند، توصیه می شود که قبل از برآورد پارامترها نقش داده های آلوده در این مسأله حذف گردد. اگر داده ها در عمل آلوده نباشد بکارگیری روش نیرومند نتایج تحلیل را معمولاً تغییر نمی دهد. به عبارت دیگر پارامترها با روش نیرومند برآورد گردند. در نظریه تصمیم انتخاب یک قاعده متضمن ضرری است که با تابع مخاطره سنجیده می‌شود و سعی بر آن است ، قاعده‌ای انتخاب شود که تابع مخاطره ی آن برای کلیه مقادیر پارامتر حداقل باشد. از آنجائی که چنین قاعده‌ای در حالت کلی وجود ندارد، سعی می‌شود قاعده‌ای را انتخاب کنند که به مفهومی بهترین باشد و یکی از این مفاهیم نااریبی است . نااریبی از قدیمی‌ترین خواصی است که انتظار می‌رود یک برآورد کننده داشته باشد. این مفهوم ابتدا توسط گاوس تحت عنوان "عدم اشتباه سیستماتیک " معرفی شد. بعدها با معرفی نظریه تصمیم مفهوم نا اریبی بر اساس تابع ضرر با نام ریسک -نااریبی توسعه یافت . همچنین یکی دیگر از مفاهیم نظریه تصمیم بیض بودن یک برآورد کننده می‌باشد. در ریسک نااریبی وبیز هدف حداقل کردن تابع مخاطره است با این تفاوت که در ریسک - نااریبی نقش برآورده کننده با پارامتر عوض شده است . حال مثلا" اگر تابع ضرر به صورت مربع اشتباهات باشد

www.matlabproject.ir

در این صورت آمارشناس می‌خواهد قاعده‌ای را انتخاب کند که فاصله برآورد کننده از برآورد شونده حداقل شود. با این تفاوت که در بیض حداقل کردن روی قواعد تصمیم و در ریسک - نااریبی روی توابع پارامتر انجام می‌گیرد. از طرفی حداقل کردن فاصله در یک فضای هیلبرت توسط روش تصویر کردن متعامد انجام می‌شود. سئوالی که پیش می‌آید این است که "آیا یک برآورد کننده بیز می‌تواند ریسک - ناریب باشد"؟ بیکل و بلکول در سال ۱۹۶۷ نشان دادند که در حالت کلی پاسخ به سئوال فوق منفی است . با استفاده از خواص تصویر کردن متعامد این مطلب را ثابت کرده و نشان داده می‌شود در بعضی حالات با استفاده از روش "جکنایف " می‌توان اریبی برآورد کننده‌های بیض را کم کرد در حالیکه برآورد کننده‌های حاصله هنوز با پیشین مناسب بیز هستند. پس از معرفی خواص تصویر کردن متعامد و توجه به این نکته که اساس آنالیز واریانس در جداول چند بعدی تجزیه بردار مشاهدات به مجموع چند بردار متعامد می‌باشد، این سئوال پیش می‌آید که "آیا همین تجزیه را می‌توان برای مسئله تصمیم انجام داد"؟ برای این منظور روشهای بیضی لگ - خطی را برای آنالیز جداول چند بعدی قابل فهم است. در روش فوق لگاریتم را به صورت ترکیب خطی از اثرات سطر، ستون و متقابل نوشته و با قراردادن پیشین روی اثرات مختلف سعی می‌شود تحت شرایطی برآوردهائی برای اثرات مختلف بدست آورد.

پس در نهایت می توان گفت استفاده از روش های پارامتریک و مخصوصا الگوی بیضین به شرط کافی بودن تعداد نمونه ها و عدم پیچیدگی زیاد ویژگی ها بهترین انتخاب است اما هر چه پیچیدگی‌ها بیشتر باشد باید به سمت استفاده تکنیکهای غیرپارامتریک برویم تا خطای تخمین کمتری داشته باشیم.

www.matlabproject.ir

منابع :

1]http://archive.ics.uci.edu/ml/
2]PR toolbox 3.2.5
3]K-Mean(EM) و بیزی ساده الگوریتمهای مقایسه و بررسی
4]Discriminative Feature Extraction and
5]Dimension Reduction
6]PCA, LDA and LSA -
7]New approaches to PCA-based eigenvoice model and LDA-based feature extraction