



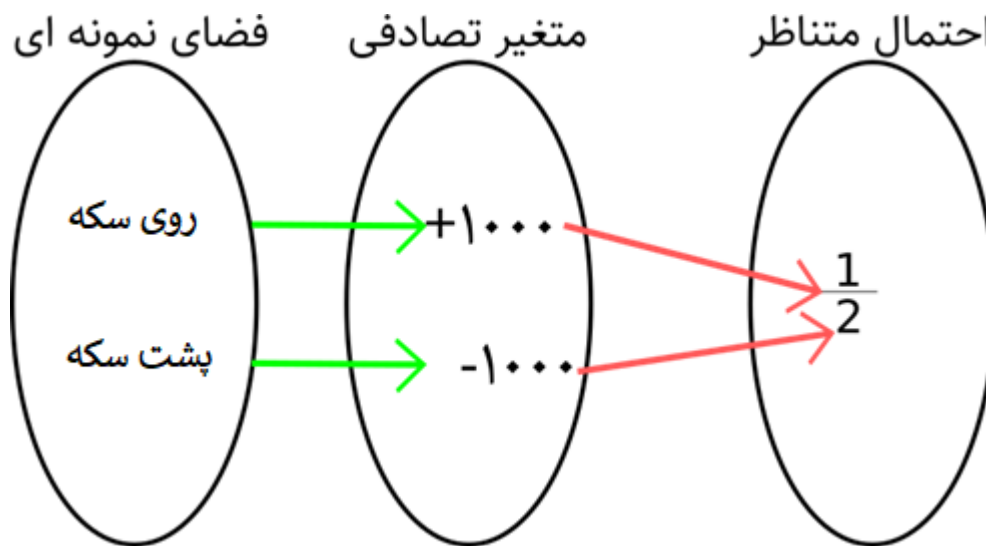
فصل اول شناسایی الگو متغیر تصادفی – تابع توزیع

محمدجواد فدائی اسلام

متغیر تصادفی

در انجام آزمایش‌ها علاقه‌مند هستیم تا خروجی تصادفی آزمایش را به صورت عددی بیان کنیم. متغیر تصادفی X تابعی است که یک عدد حقیقی را به یک خروجی نگاشت می‌کند.

این تابع به دو صورت گسسته و پیوسته در نظر گرفته می‌شود.



تابع توزیع تجمعی = تابع توزیع احتمال

○ تابع توزیع تجمعی یک متغیر تصادفی، تابعی است که **احتمال** **رخدادن پیشامدهای با مقدار عددی کمتر از آن** را نمایش می‌دهد. و به صورت دقیق به شکل زیر تعریف می‌شود:

$$F(x) = P[X \leq x]$$

- Cumulative Distribution Function (CDF)
- Probability Distribution Function (PDF)

تابع توزیع تجمعی

○ تابع توزیع تجمعی متغیرهای تصادفی از نوع گسسته و پیوسته به ترتیب به صورت زیر تعریف می شوند.

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$$

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$$

خواص تابع توزیع تجمعی (گسسته یا پیوسته)

$$۱- \quad 0 \leq P(X \leq x) \leq 1 \quad \text{یا} \quad 0 \leq F(x) \leq 1$$

۲- $F(x)$ یک تابع غیر نزولی است.

$$۳- \quad \lim_{x \rightarrow -\infty} F(x) = F(-\infty) = 0 \quad \text{و} \quad \lim_{x \rightarrow +\infty} F(x) = F(+\infty) = 1$$

۴- $F(x)$ در هر نقطه x از راست پیوسته است.

خواص تابع توزیع تجمعی (ادامه)

$$P(a \leq X \leq b) = F(b) - F(a) \quad -5$$

۶- الف: در متغیر پیوسته

$$f(x) = \frac{d F(x)}{dx}$$

$$F(x) = \int_{-\infty}^x f(t) dt$$

ب: در متغیر گسسته

$$f(x) = F(x) - F(x^-) \quad F(x) = \sum_{x \leq t} f(t)$$

تابع چگالی احتمال پیوسته

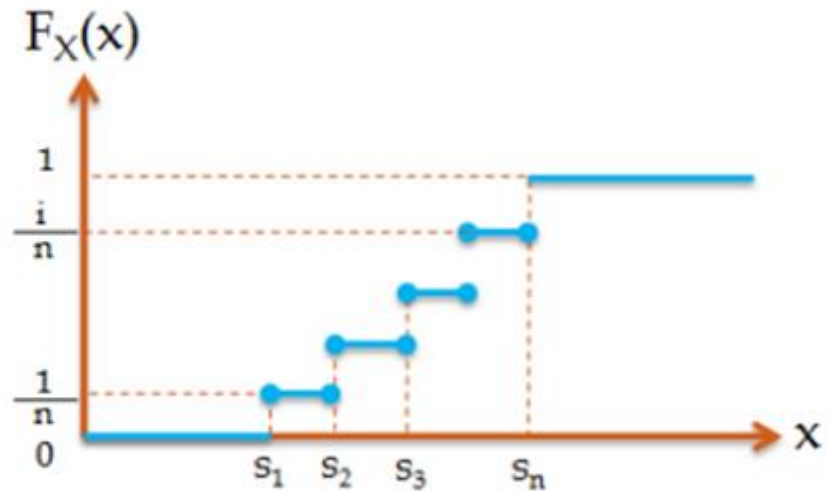
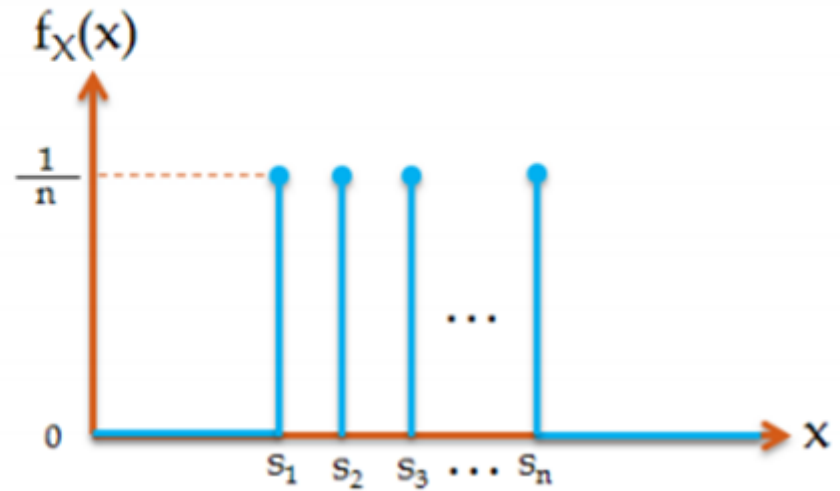
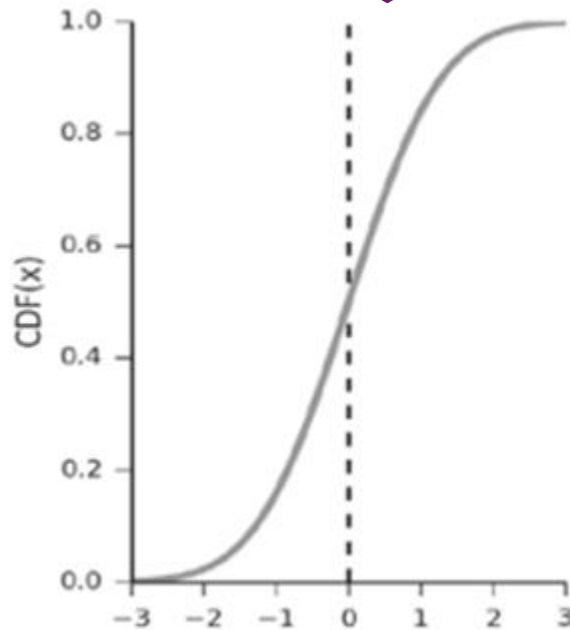
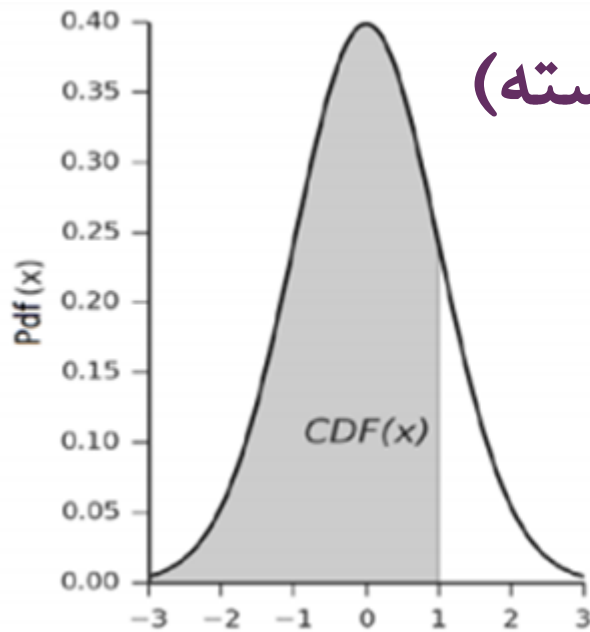
PROBABILITY DENSITY FUNCTION(PDF)

○ مشتق تابع توزیع تجمعی، تابع چگالی احتمال نام دارد.

$$f(x) = \frac{dF(x)}{dx}$$

○ تابع چگالی احتمال پیوسته، **احتمالی را نشان نمی‌دهد** و برای به دست آوردن احتمال باید از آن انتگرال گرفت.

تابع چگالی و تابع توزیع (پیوسته و گسسته)



امید ریاضی

EXPECTED VALUE

اگر متغیر تصادفی X دارای تابع چگالی احتمال $f(x)$ باشد. امید ریاضی در حالت گسسته و پیوسته به ترتیب به صورت زیر تعریف می شود.

$$E[X] = \sum_x x \cdot f(x)$$

$$E[X] = \mu = \int_{-\infty}^{\infty} x f(x) dx$$

در ادبیات آماری امید ریاضی را معمولاً با μ نمایش می دهند. امید ریاضی برابر است با مقداری که به طور متوسط از یک فرایند تصادفی با بی نهایت تکرار انتظار می رود.

ویژگی‌های امید ریاضی

$$E(X + c) = E(X) + c$$

$$E(X + Y) = E(X) + E(Y)$$

$$E(aX) = a E(X)$$

$$E(aX + b) = a E(X) + b$$

$$E(aX + bY) = a E(X) + b E(Y)$$

اگر X و Y مستقل از هم باشند.

$$E(XY) = E(X).E(Y)$$

تفاوت امید ریاضی و میانگین

- این دو عبارت گاهی به جای یکدیگر در آمار استفاده می‌شوند.
- زمانی که بخواهیم **میانگین یک توزیع احتمال** را محاسبه کنیم از مقدار عبارت امید ریاضی استفاده می‌شود. این نشان‌دهنده میانگین مقداری است که انتظار داریم قبل از جمع‌آوری داده‌ها رخ دهد.
- میانگین معمولاً زمانی استفاده می‌شود که بخواهیم مقدار متوسط تعدادی نمونه **جمع‌آوری** شده را محاسبه کنیم.

Goals (X)	Probability P(X)
0	0.18
1	0.34
2	0.35
3	0.11
4	0.02

تفاوت امید ریاضی و میانگین - مثال

○ جدول روبرو، تابع توزیع احتمال تعداد گل‌هایی است که یک تیم معین در یک بازی به ثمر می‌رساند.

$$\mu = \sum xP(x) = 1.45$$

○ تعداد گل‌هایی که یک تیم در ۱۵ بازی زده است به صورت زیر است

1, 1, 0, 2, 2, 1, 0, 3, 1, 1, 1, 2, 4, 3, 1

$$Mean = \sum \frac{x_i}{n} = 1.53$$

باز تعریف واریانس از روی تابع چگالی احتمال

اگر تابع چگالی احتمال $f(x)$ در دامنه $[a, b]$ در دسترس باشد واریانس آن به صورت زیر قابل محاسبه است:

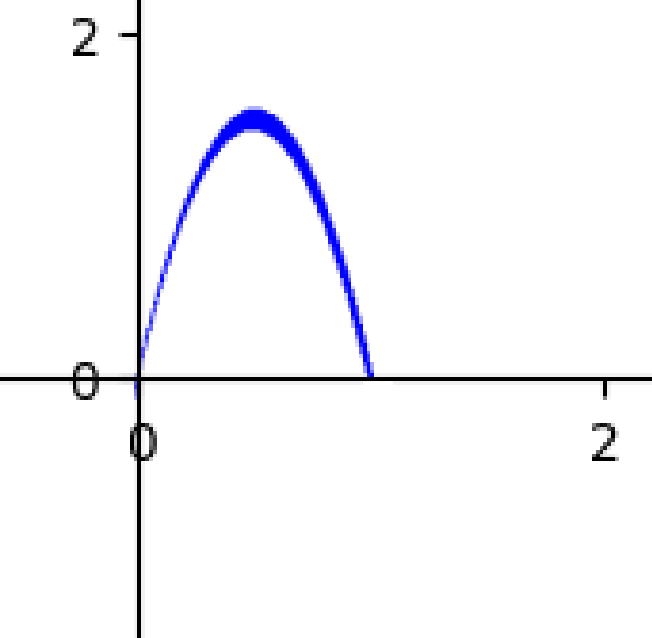
$$V(x) = \int_a^b (x - \mu)^2 f(x) dx = \sigma^2 = E[(x - \mu)^2]$$

باز تعریف واریانس از روی تابع چگالی احتمال

باتوجه به خواص عملگر E می توان واریانس X را به صورت زیر تعریف کرد.

$$\begin{aligned} V(X) &= E[(X - \mu)^2] = E[X^2 + \mu^2 - 2\mu X] \\ &= E(X^2) + \mu^2 - 2\mu E(X) \\ &= E(X^2) + \mu^2 - 2\mu^2 = E(X^2) - \mu^2 \end{aligned}$$

مثال



○ واریانس تابع زیر را حساب کنید

○ $f(x) = 6x - 6x^2, 0 \leq x \leq 1$

○ $\mu = \int_0^1 x(6x - 6x^2)dx = \frac{1}{2}$

○ $V(x) = \int_0^1 \left(x - \frac{1}{2}\right)^2 (6x - 6x^2)dx = \frac{1}{20}$

ویژگی های واریانس

$$V(c) = 0$$

$$V(aX) = a^2 V(X)$$

$$V(aX + c) = a^2 V(X) + 0$$

$$\begin{aligned} V(aX + c) &= E[aX + c - E(aX + c)]^2 \\ &= E[aX + c - aE(X) - c]^2 = E[aX - aE(X)]^2 \\ &= E[a(X - E(X))]^2 = a^2 E(X - \mu)^2 = a^2 V(X) \end{aligned}$$

کوواریانس دو متغیر تصادفی

اگر متغیرهای تصادفی X و Y دارای تابع چگالی احتمال توام $f(x,y)$ باشند کوواریانس آنها به صورت زیر تعریف می شود.

$$\text{Cov}(X, Y) = E[(X - \mu_x)(Y - \mu_y)]$$

که μ_x و μ_y به ترتیب امید ریاضی X و Y هستند. رابطه بالا را می توان به صورت زیر نیز نوشت:

$$\text{Cov}(X, Y) = E[XY] - \mu_x \mu_y$$

قضیه: اگر دو متغیر X و Y مستقل باشند آنگاه:

$$\text{Cov}(X, Y) = 0$$

کواریانس نمونه‌ها

○ اگر N نمونه تصادفی در دست باشد که از هر نمونه دو ویژگی X و Y در دست باشد. ماتریس کواریانس این دو ویژگی یک ماتریس دو در دو به صورت زیر خواهد بود.

$$\begin{bmatrix} cov(X, X) & cov(X, Y) \\ cov(Y, X) & cov(Y, Y) \end{bmatrix}$$

$$cov(X, X) = var(X), cov(Y, Y) = var(Y)$$

$$cov(X, Y) = cov(Y, X) = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})$$

$$\text{cov}(X, a) = 0$$

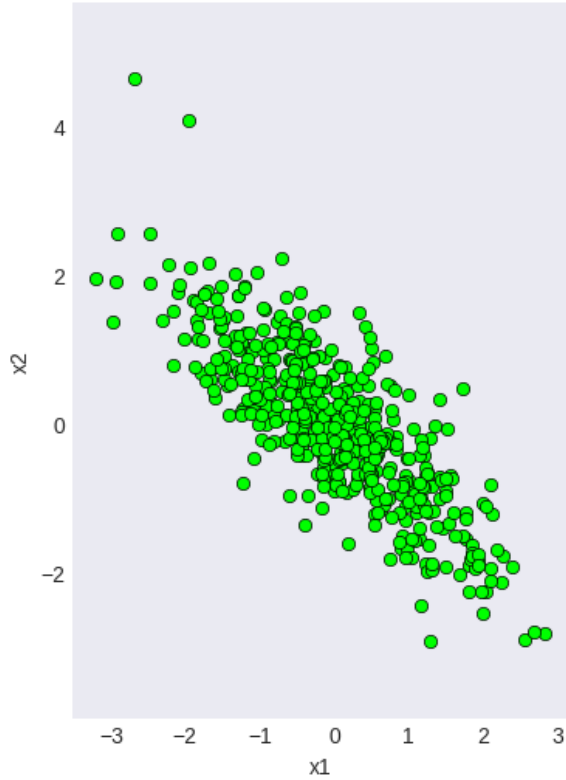
$$\text{cov}(X, X) = \text{var}(X)$$

$$\text{cov}(X, Y) = \text{cov}(Y, X)$$

$$\text{cov}(aX, bY) = ab \text{ cov}(X, Y)$$

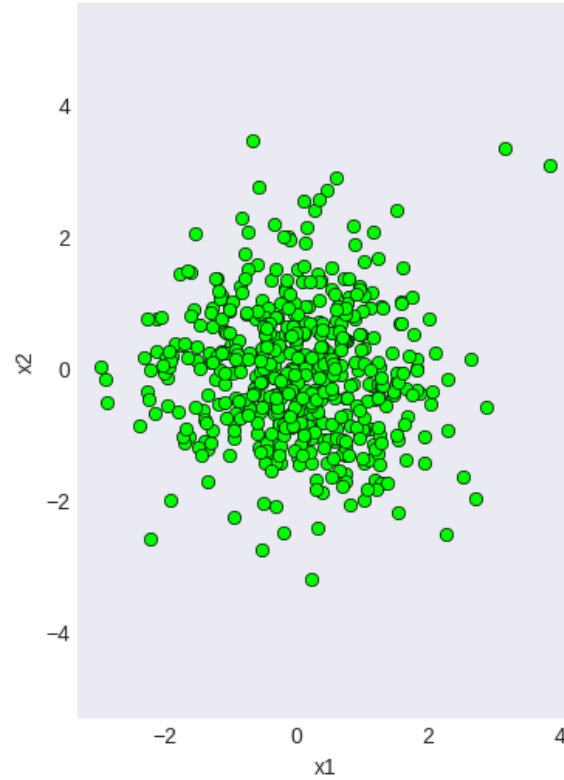
$$\text{cov}(X + a, Y + b) = \text{cov}(X, Y)$$

Covariance between x1 and x2 = -0.8



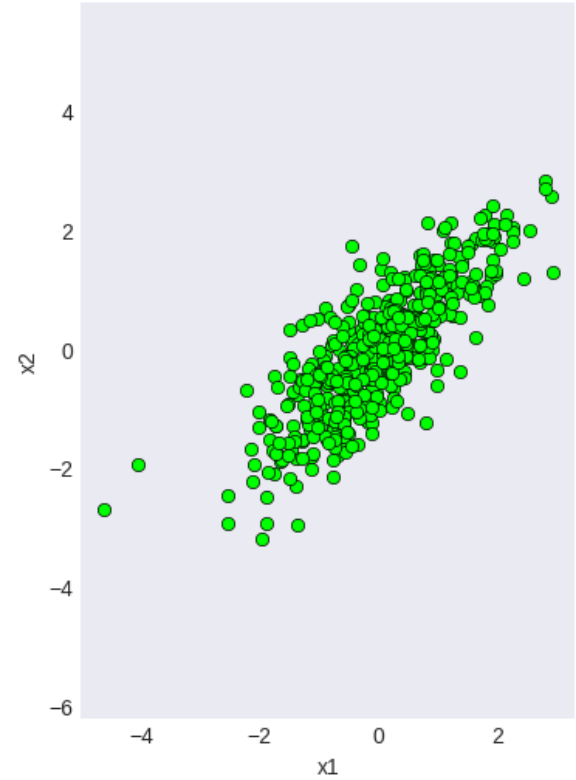
$$cov = \begin{bmatrix} 1 & -0.8 \\ -0.8 & 1 \end{bmatrix}$$

Covariance between x1 and x2 = 0



$$cov = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Covariance between x1 and x2 = 0.8



$$cov = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$$

COVARIANCE

- A positive covariance between two variables reveals that the paired values of both variables tend to increase together. A negative covariance reveals that there is an inverse relationship between the variables, that is, as one increases, the other tends to decrease.
- Weight loss reduces diabetes risk. Lack of exercise increases the risk of diabetes.
- $\text{cov}(\text{weight}, \text{diabetes risk}) > 0$
- $\text{cov}(\text{sport}, \text{diabetes risk}) < 0$
- $\text{cov}(\text{sport}, \text{HIV}) = 0$



توزیع یکنواخت، توزیع نرمال

محمد جواد فدائی اسلام

تابع احتمال یکنواخت گسسته

- متغیر تصادفی X دارای تابع احتمال یکنواخت با پارامتر k است اگر تابع احتمال آن به صورت زیر باشد:

$$f(x) = P(X = x) = \frac{1}{k} \quad x = 1, 2, \dots, k$$

- مثال: کیسه‌ای شامل k گوی یکسان است که بر روی هر کدام شماره‌ای از ۱ تا k وجود دارد. احتمال برداشتن گوی به شماره i چقدر است. i بین یک و k است.

تابع احتمال یکنواخت گسسته

مثال: جعبه‌ای شامل کلیدها با شماره ۱ تا k است. اگر هم‌شانس بودن را برای همه شماره‌ها در نظر بگیریم و متغیر تصادفی X ، شماره کلید خارج شده باشد، آنگاه X دارای تابع چگالی احتمال یکنواخت گسسته است.

$$E(X) = \sum_{x=1}^k x f(x) = \sum_{x=1}^k x \cdot \frac{1}{k} = \frac{1}{k} \sum_{x=1}^k x = \frac{k+1}{2}$$

$$E(X^2) = \sum_{x=1}^k x^2 f(x) = \frac{1}{k} \sum_{x=1}^k x^2 = \frac{1}{k} \frac{k(k+1)(2k+1)}{6} = \frac{(k+1)(2k+1)}{6}$$

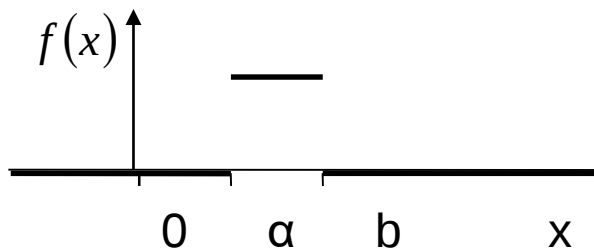
$$V(X) = E(X^2) - (E(X))^2 = \frac{(k+1)(2k+1)}{6} - \frac{(k+1)^2}{4} = \frac{(k+1)(k-1)}{12}$$

تابع چگالی احتمال یکنواخت (مستطیلی)

UNIFORMLY PROBABILITY DENSITY FUNCTION

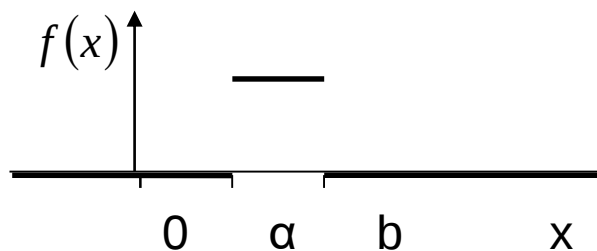
○ متغیر تصادفی X دارای تابع چگالی احتمال یکنواخت با پارامترهای a و b است اگر تابع چگالی آن به صورت زیر باشد.

$$f(x) = \frac{1}{b-a} \quad a < x < b$$



ویژگی‌های تابع چگالی احتمال یکنواخت

○ نمودار چگالی احتمال $f(X)$ برای $-\infty < a < b < \infty$ به صورت زیر است.



○ تابع توزیع $F(X)$ برابر است با:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x < b \\ 1 & x \geq b \end{cases}$$

تابع چگالی احتمال نرمال

NORMAL (GAUSSIAN) PROBABILITY DENSITY FUNCTION

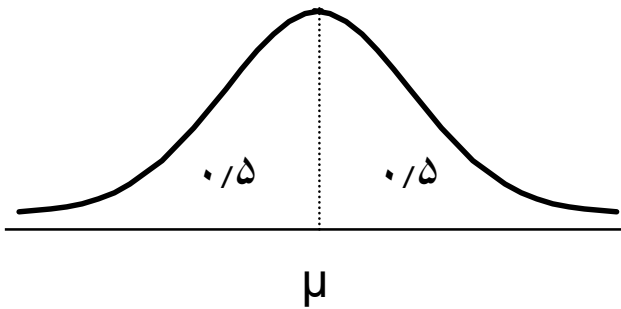
○ متغیر تصادفی نرمال یکی از توزیع‌های مهم آماری در حالت پیوسته است. متغیر تصادفی X دارای توزیع نرمال با میانگین μ و واریانس σ^2 است، اگر تابع چگالی احتمال آن به صورت زیر باشد.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < +\infty$$

ویژگی‌های توزیع نرمال

۱- این توزیع نسبت به محور $y=\mu$ دارای تقارن است.

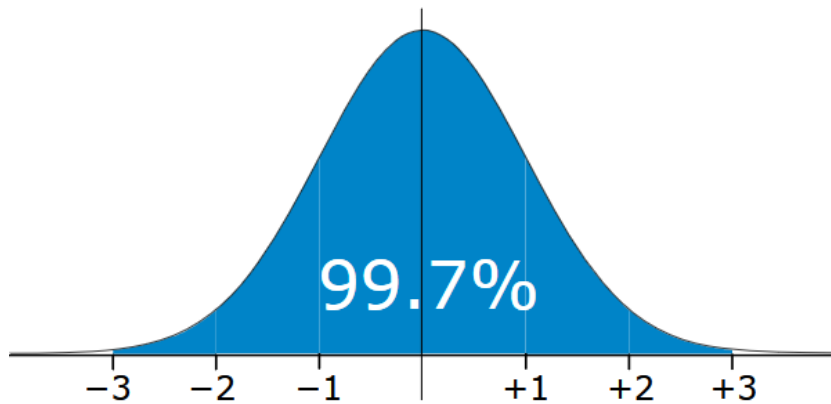
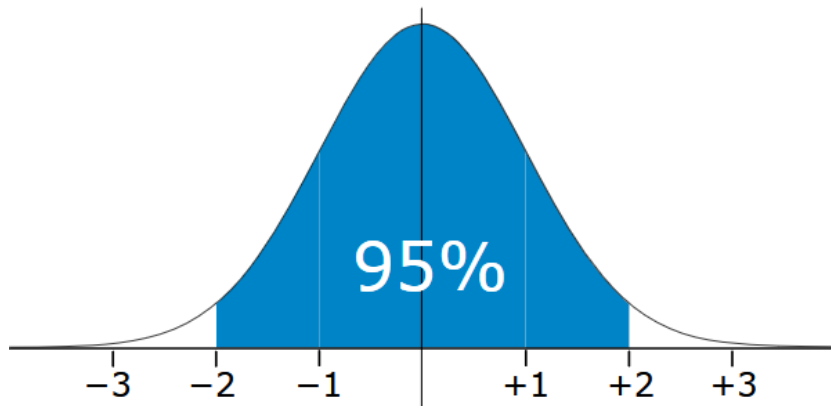
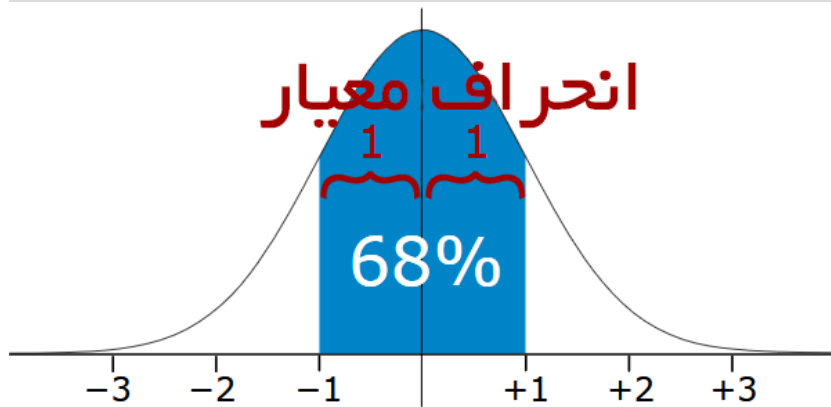
۲- $\int_{-\infty}^{\infty} f(x) dx = 1$



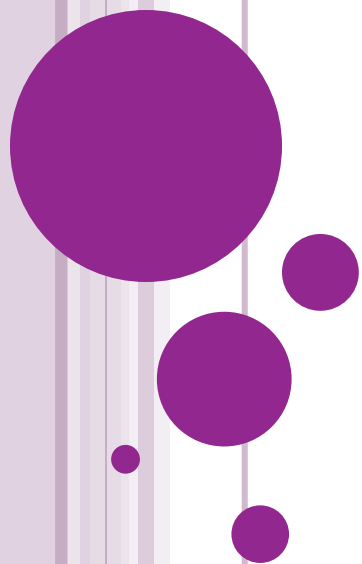
۳- $P[X \leq \mu] = P[X \geq \mu] = 0.5$

۴- برای $\mu=0$ و $\sigma^2 = 1$ ، توزیع نرمال را توزیع نرمال استاندارد گویند.

توزیع نرمال (ویژگی)



- تصویر اول: ۶۸٪ از مقادارها در محدوده یک انحراف از میانگین
 - تصویر دوم: ۹۵٪ مقادارها در محدوده دو برابر انحراف معیار از میانگین
 - تصویر سوم: ۹۹.۷٪ از مقادارها در محدوده سه برابر انحراف معیار از میانگین.
- (انحراف معیار در اشکال روبرو برابر یک در نظر گرفته شده است.)



ابزار MATLAB در رسم نمودارها

رسم تابع با استفاده از دستور EZPLOT (رسم با معادله و سمبول)

% Easy-to-use function plotter

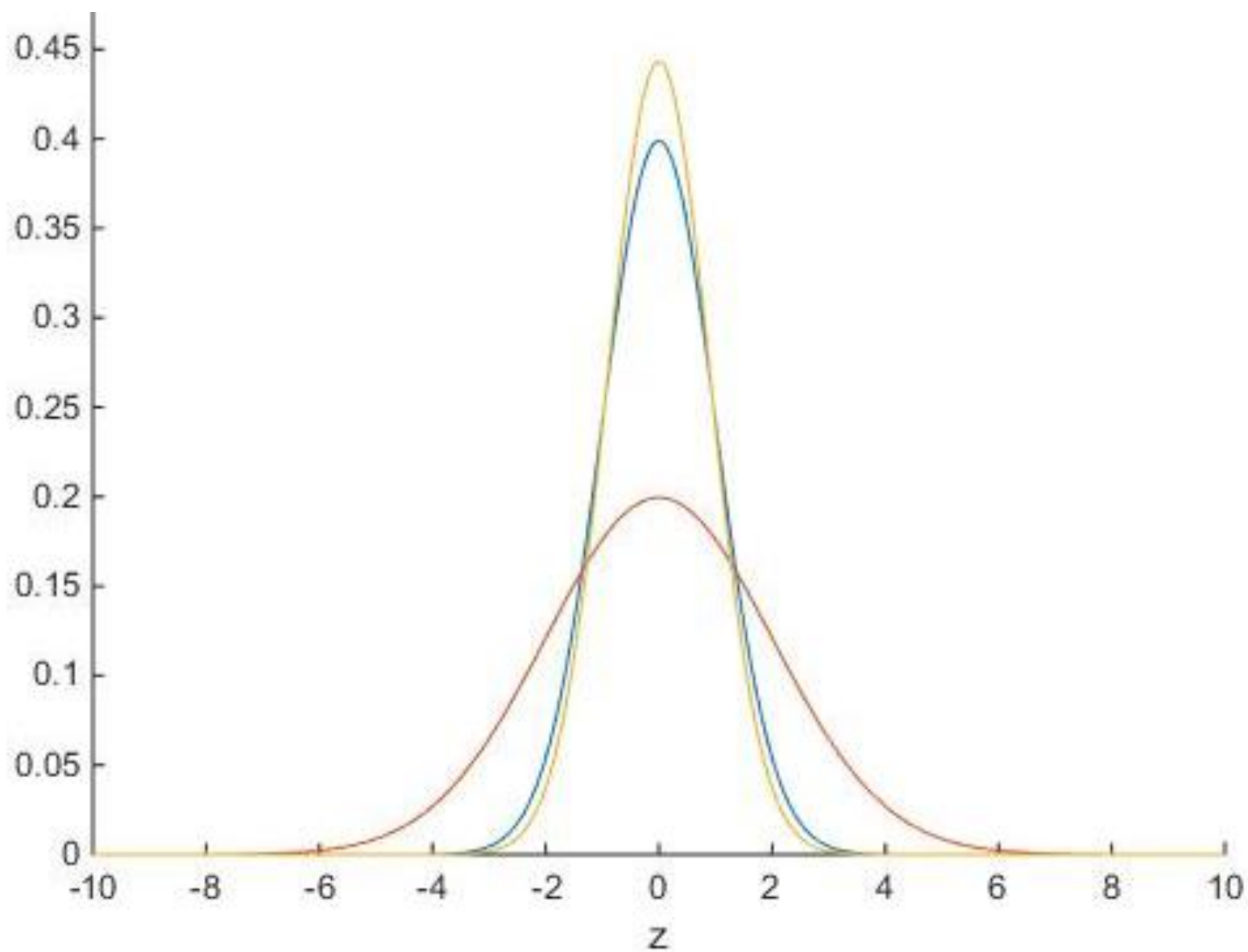
figure; hold on

ezplot('(2*pi)^-0.5 * exp(-1/2*(z^2))',[-10 10 0 0.5]);

ezplot('(2*pi*4)^-0.5 * exp(-1/2*(z/2)^2)',[-10 10 0 0.5]);

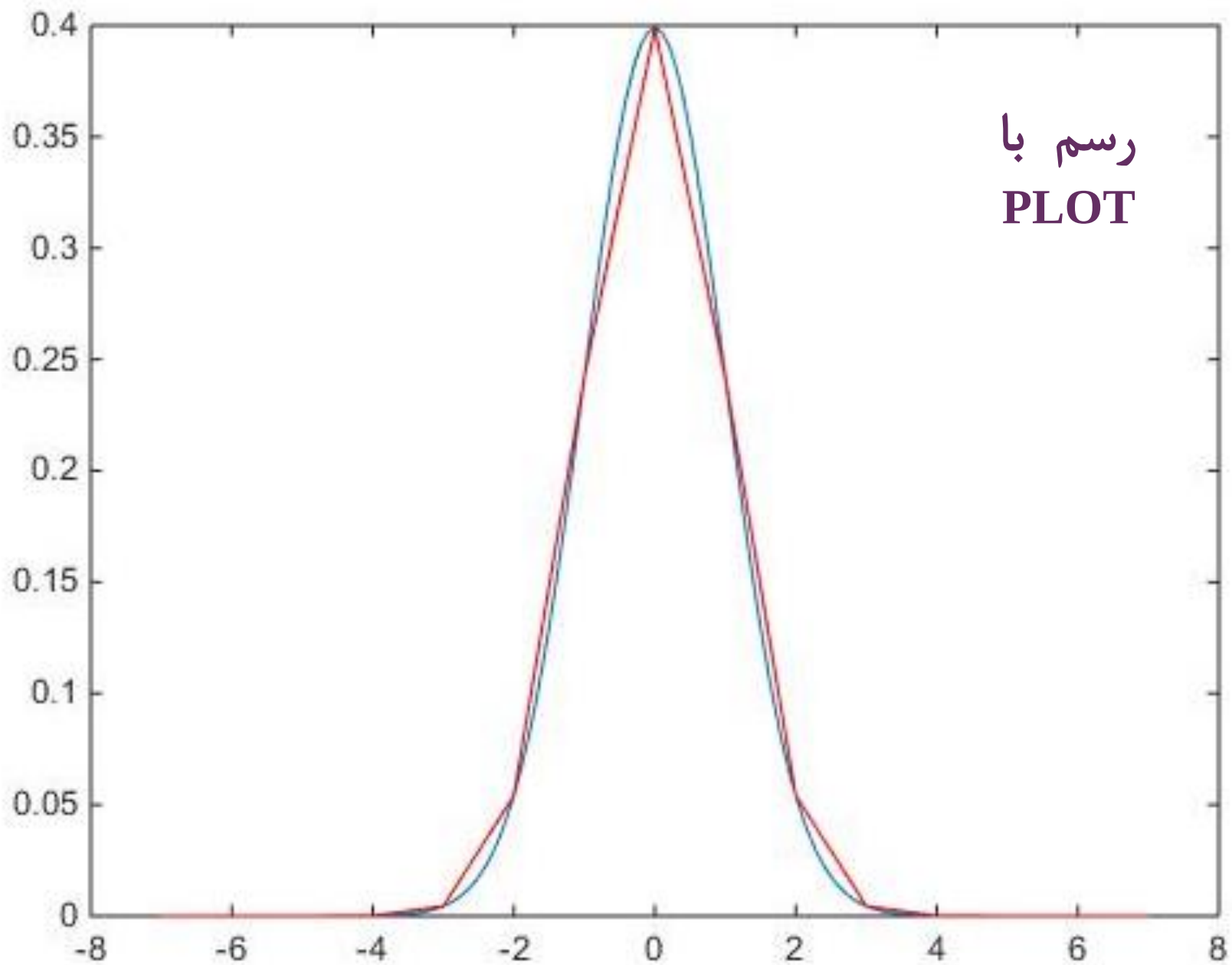
ezplot('(2*pi*0.81)^-0.5 * exp(-1/2*(z/.9)^2)',[-10 10 0 0.5]);

رسم با EZPLOT



رسم تابع با استفاده از نقطه‌یابی و دستور PLOT

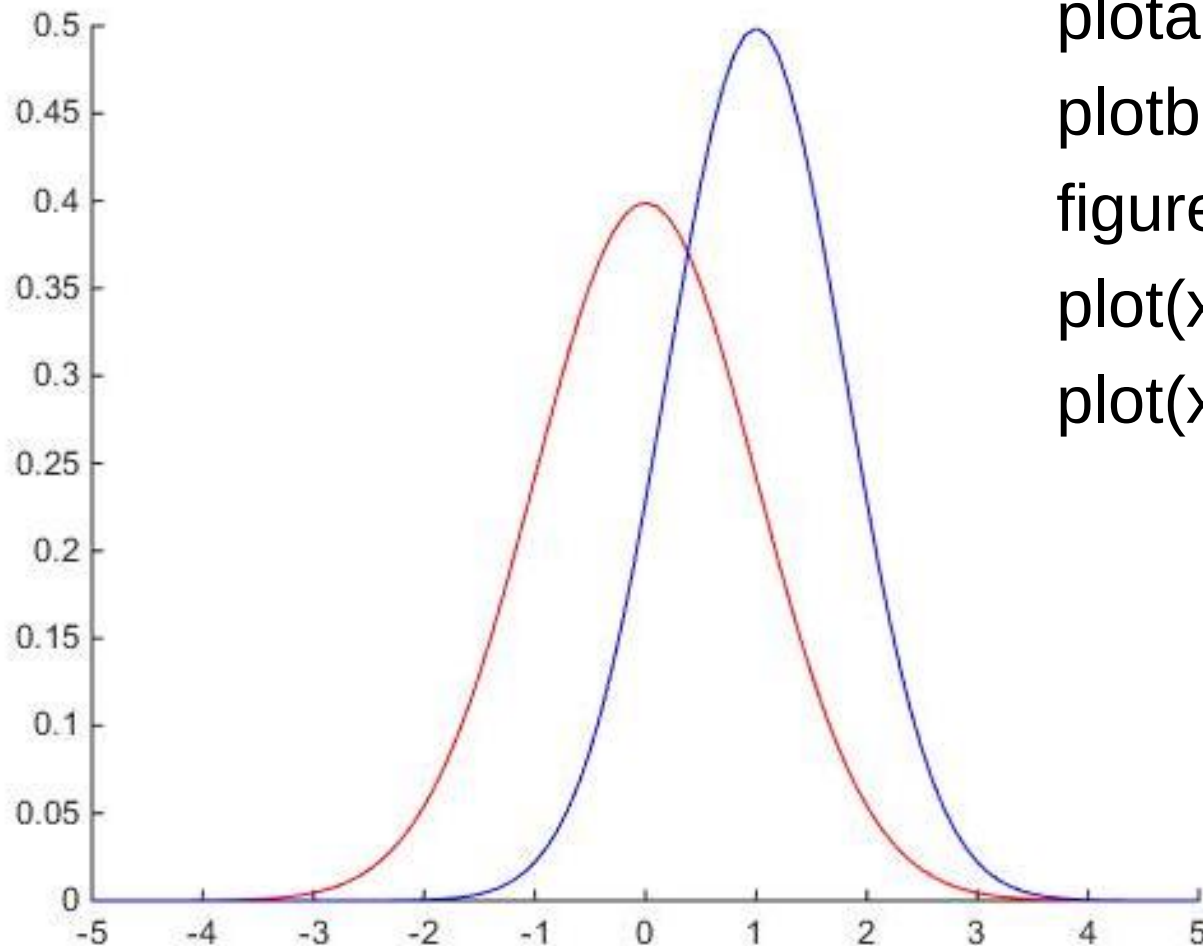
```
% plot  
X = [-7:1:7];  
Y = (2*pi)^-0.5 .* exp(-1/2.*(X.^2));  
figure; plot(X,Y);  
XX = [-7:1:7];  
YY = (2*pi)^-0.5 .* exp(-1/2.*(XX.^2));  
hold on; plot(XX,YY,'r');
```



رسم با
PLOT

محاسبه تابع چگالی نرمال با NRMPDF و رسم با PLOT

```
x = [-5:.1:5];  
plota = normpdf(x,0,1);  
plotb = normpdf(x,1,0.8);  
figure; hold on;  
plot(x,plota,'r');  
plot(x,plotb,'b');
```



تابع چگالی احتمال گوسی چند متغیره

$$p(\underline{x}) = \frac{1}{(2\pi)^{\frac{\ell}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (\underline{x} - \underline{\mu})^T \Sigma^{-1} (\underline{x} - \underline{\mu})\right)$$

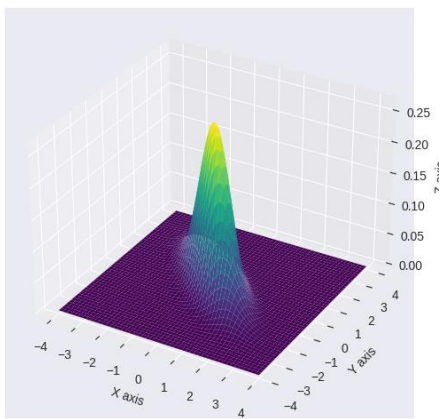
خط زیر به معنای بردار است.

Σ = covariance matrix

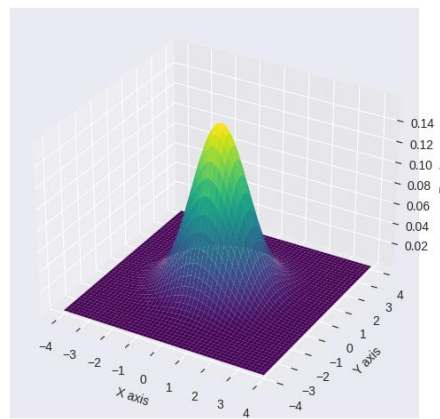
μ = mean vector

$|\cdot|$ = determinant

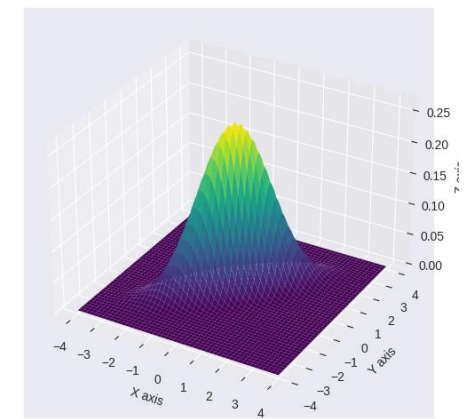
l = number of features (variables)



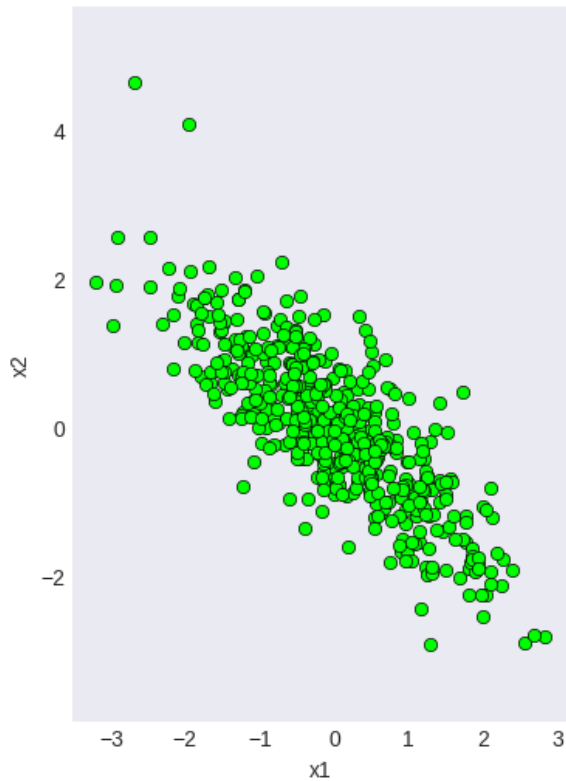
Covariance between x_1 and $x_2 = -0.8$



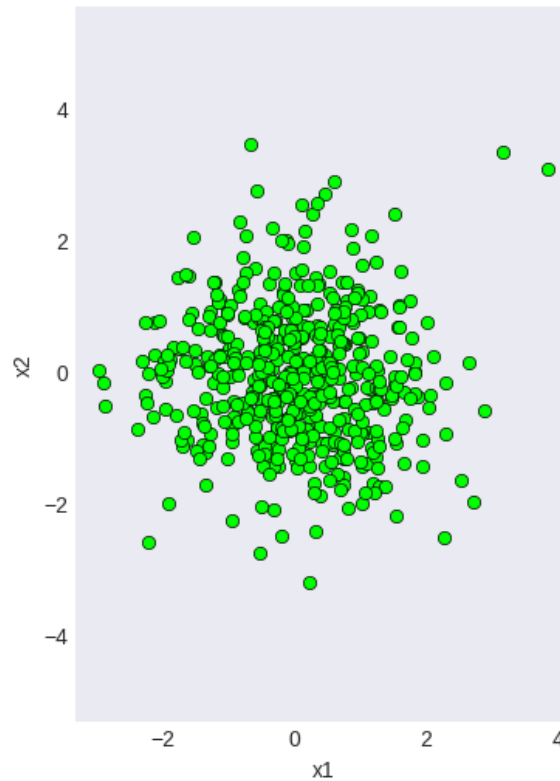
Covariance between x_1 and $x_2 = 0$



Covariance between x_1 and $x_2 = 0.8$

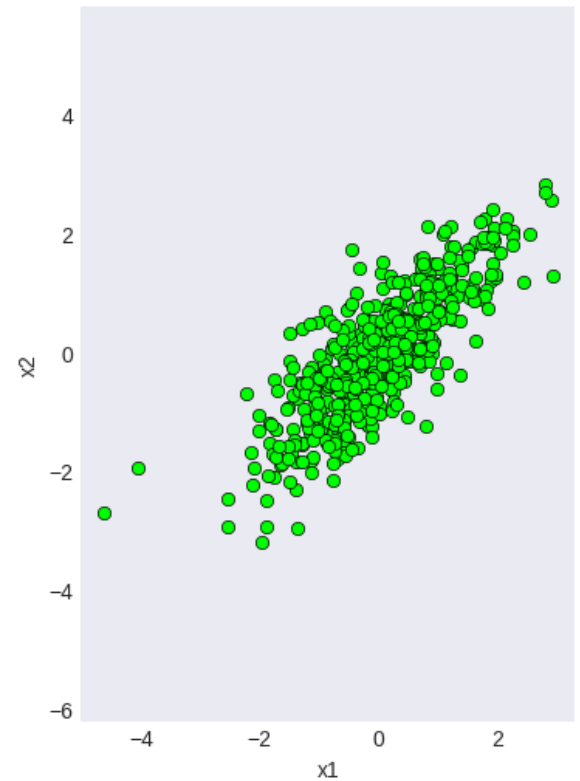


$$\text{cov} = \begin{bmatrix} 1 & -0.8 \\ -0.8 & 1 \end{bmatrix}$$



$$\text{cov} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

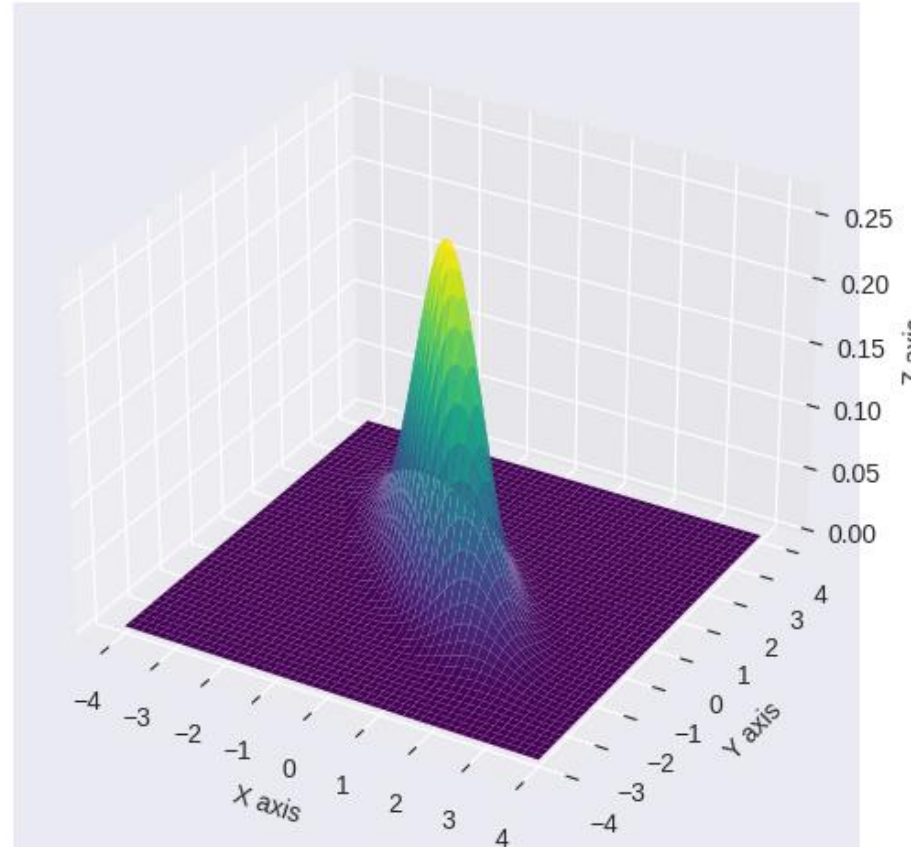
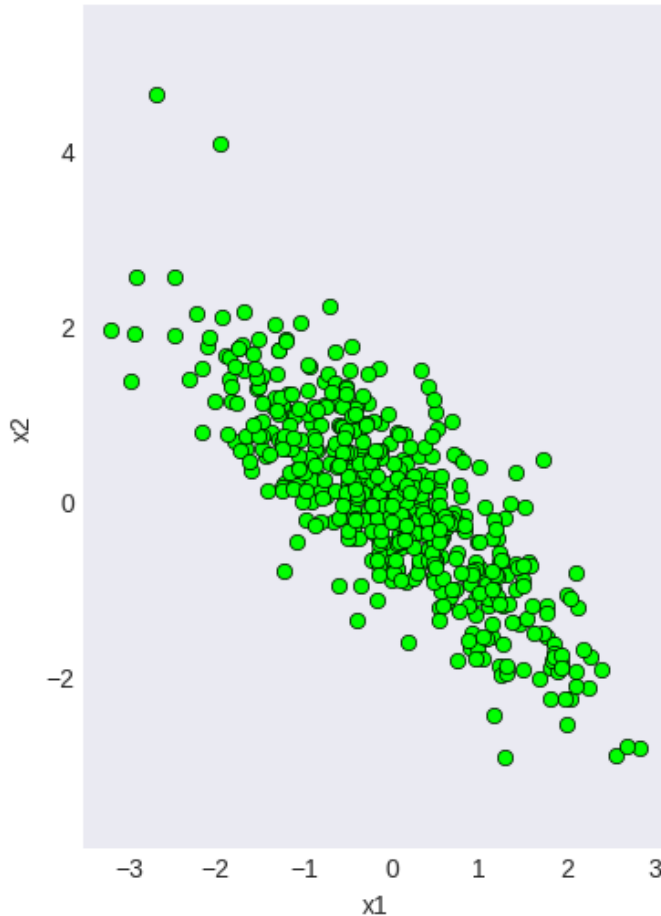
Pattern Recognition



$$\text{cov} = \begin{bmatrix} 1 & 0.8 \\ 0.8 & 1 \end{bmatrix}$$

داده و تابع چگالی توزیع متناظر آن

Covariance between x1 and x2 = -0.8

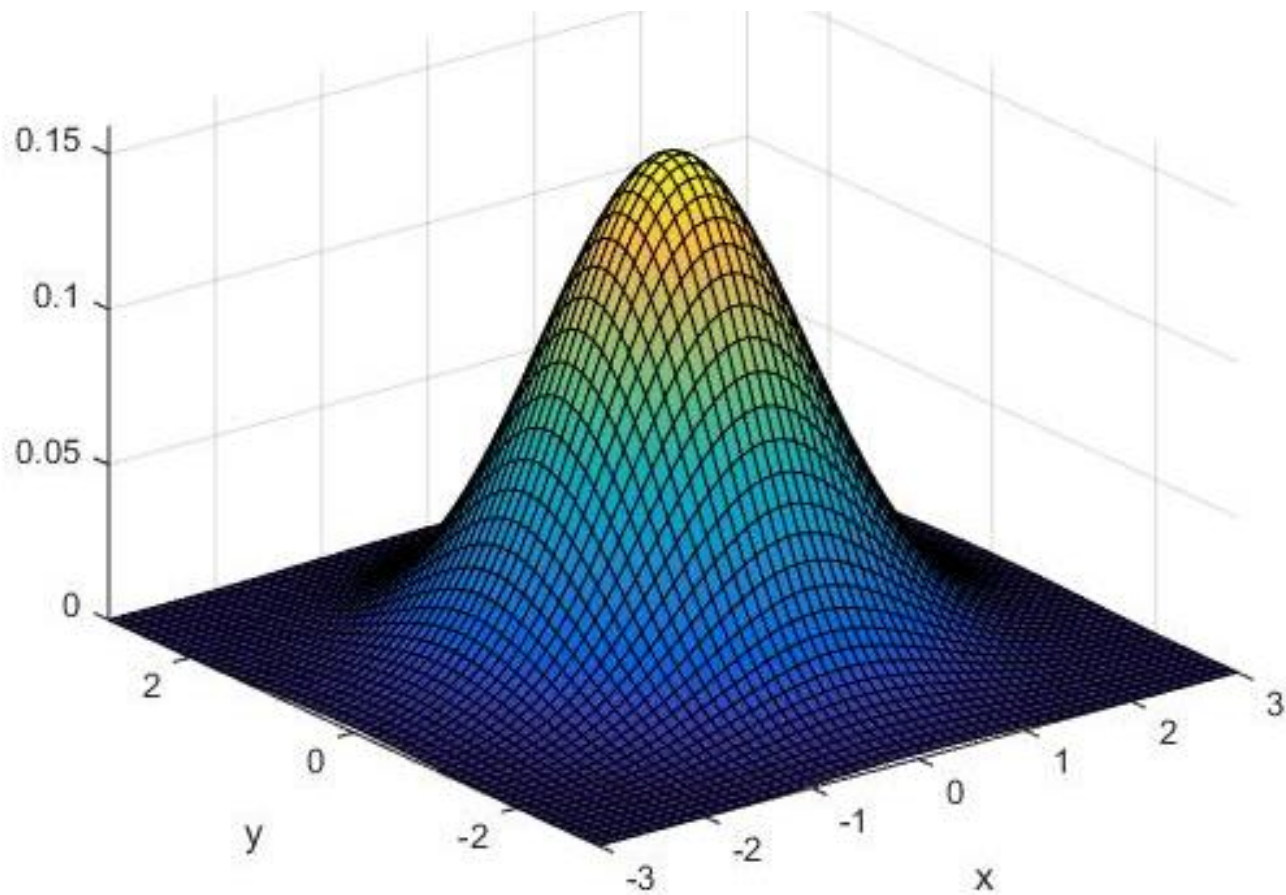


$$\text{cov} = \begin{bmatrix} 1 & -0.8 \\ -0.8 & 1 \end{bmatrix}$$

رسم رویه با استفاده از دستور EZSURF

```
ezsurf('(2*pi)^(-1)*exp((-1/2)*(x^2+y^2))',[-3,3 -3,3])
```

$$\Sigma = \begin{bmatrix} 1 & \\ & 1 \end{bmatrix}, \mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$



MESHGRID دستور

```
[X,Y] = meshgrid(1:3,10:14)
```

X =

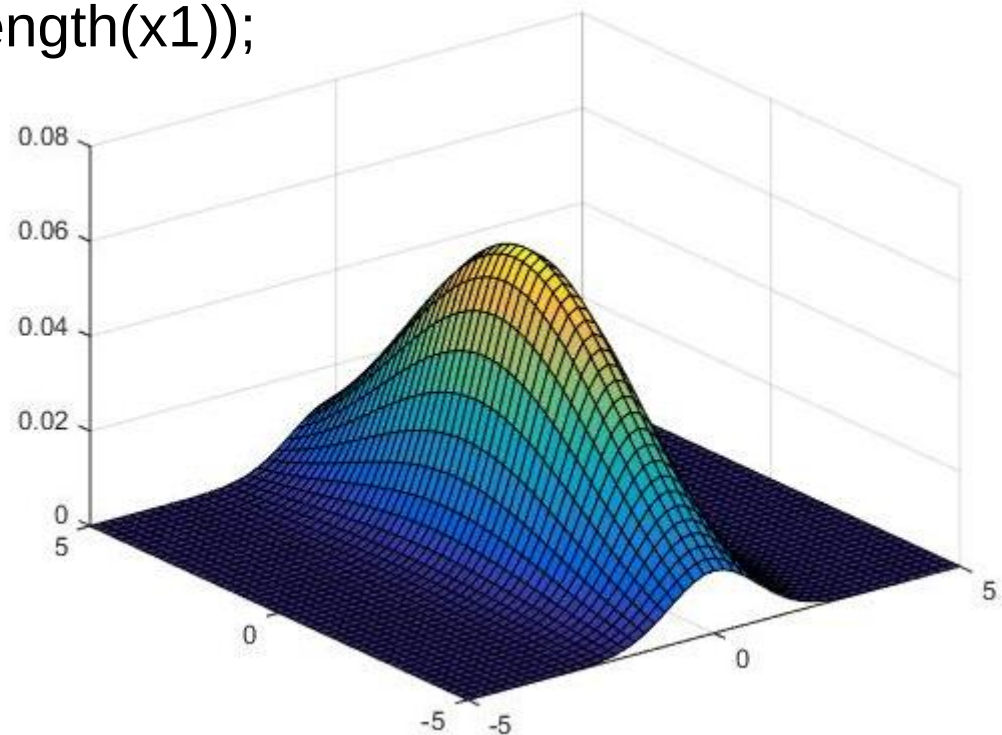
1	2	3
1	2	3
1	2	3
1	2	3
1	2	3

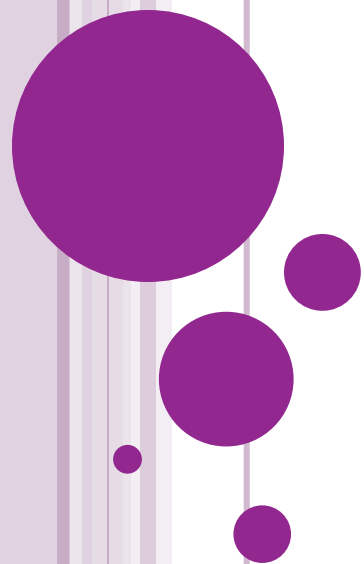
Y =

10	10	10
11	11	11
12	12	12
13	13	13
14	14	14

محاسبه تابع چگالی نرمال چند متغیره با MVNPDF و رسم رویه با PLOT

```
x1 = -5:.2:5; x2 = -5:.2:5;  
[X1,X2] = meshgrid(x1,x2);  
mu1 = [0 0];  
Sigma1 = [.8 0; 0 8];  
F1 = mvnpdf([X1(:) X2(:)],mu1,Sigma1);  
F1 = reshape(F1,length(x2),length(x1));  
%c = F1;  
%c(:) = 255;  
% surf(x1,x2,F1,c);  
surf(x1,x2,F1);
```





ابزار PYTHON در رسم نمودارها

رسم تابع با استفاده از SYMPY

(رسم با استفاده از معادله و سمبول)

```
#Python library for symbolic mathematics
```

```
from sympy.plotting import plot_implicit
```

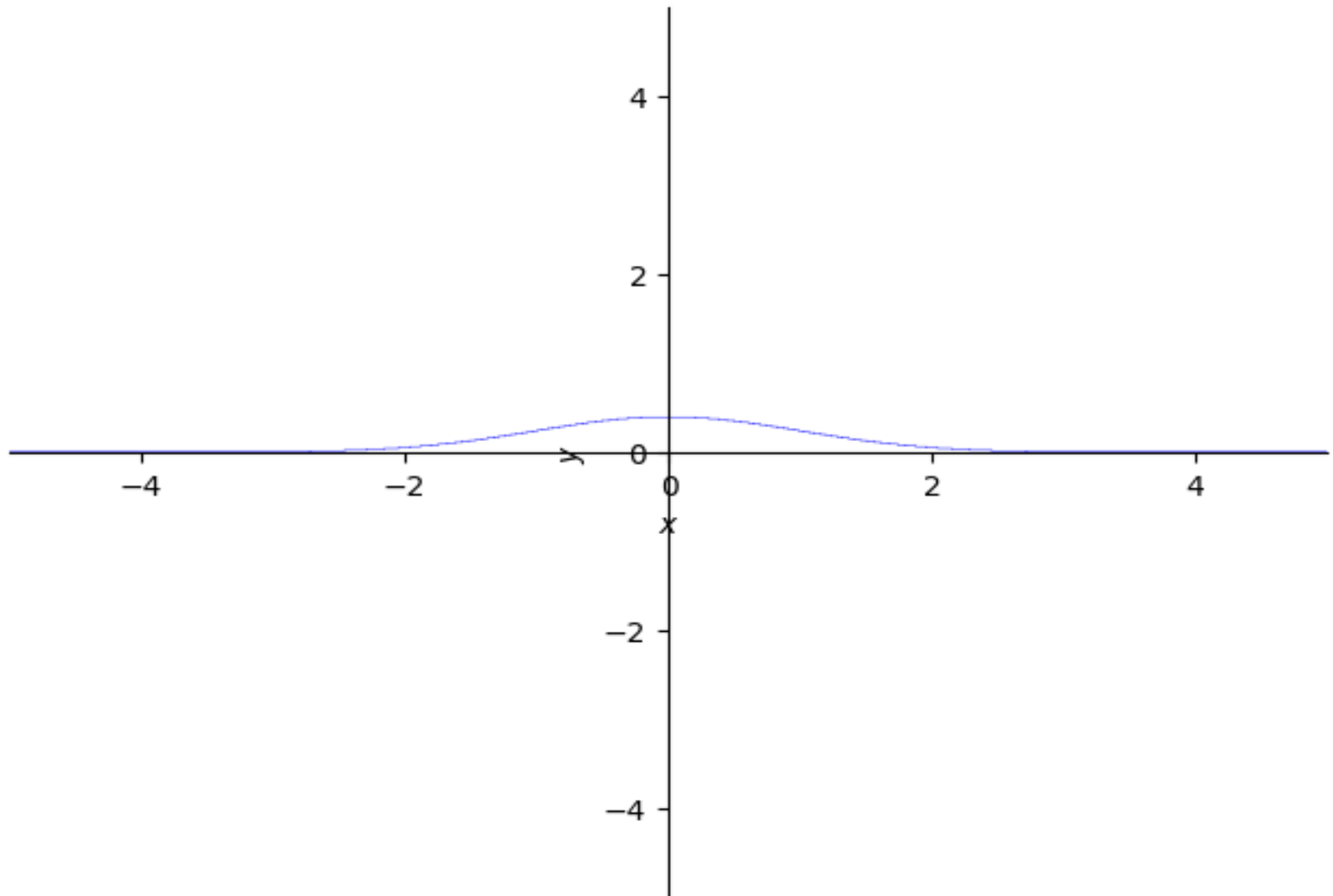
```
from sympy.parsing.sympy_parser import parse_expr
```

```
s = 'y - (2 * 3.14)**(-0.5)*exp((-1/2)*(x**2))'
```

```
eqn = parse_expr(s)
```

```
plot_implicit(eqn)
```

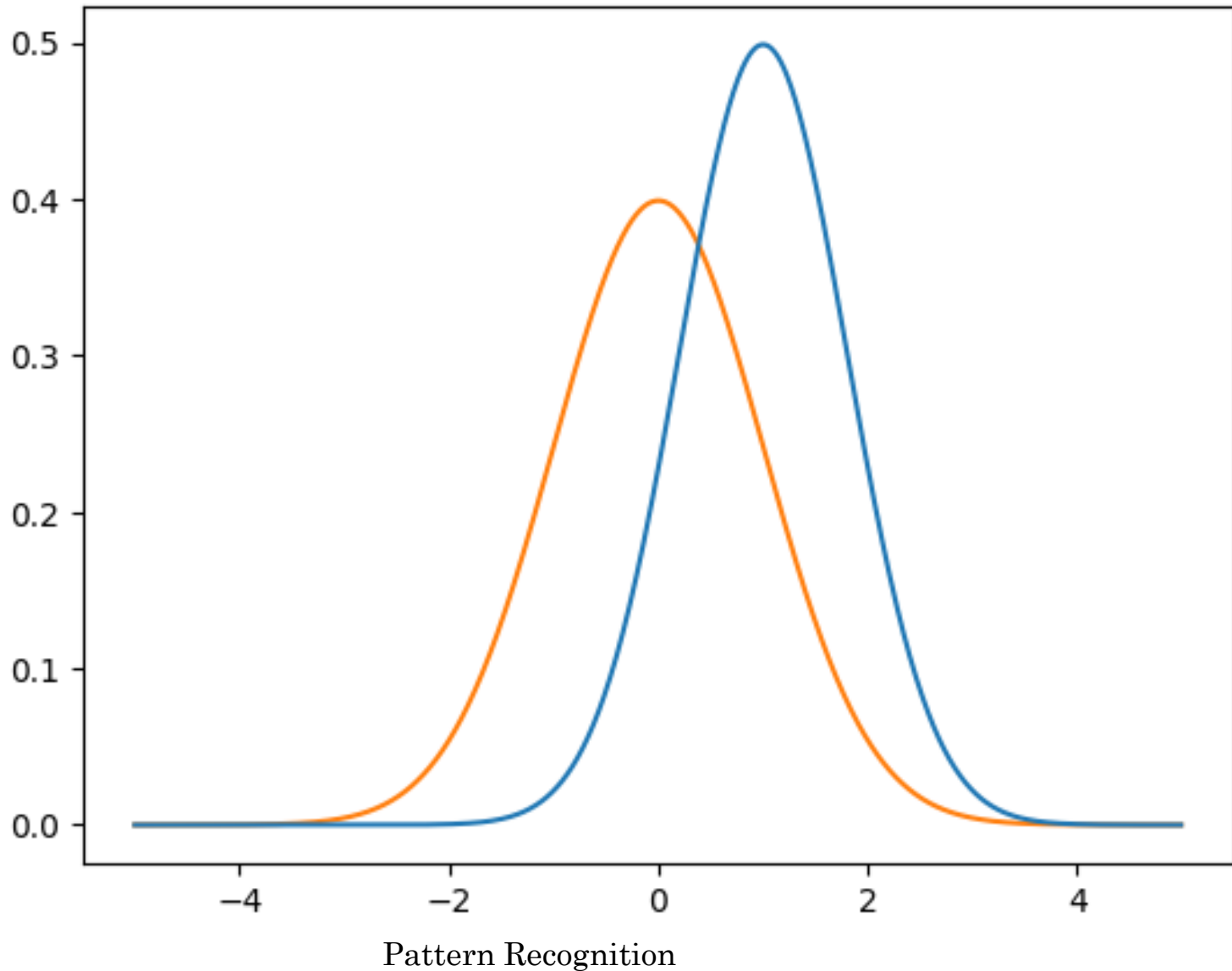
رسم تابع با استفاده از SYMPY



محاسبه تابع چگالی نرمال با NORM و رسم با PYPLOT

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm
# Plot between -5 and 5 with .001 steps.
x_axis = np.arange(-5, 5, 0.01)
mean = 0
sd = 1
plt.plot(x_axis, norm.pdf(x_axis, mean, sd), 'C1')
mean = 1
sd = 0.8
plt.plot(x_axis, norm.pdf(x_axis, mean, sd))
plt.show()
```

محاسبه تابع چگالی نرمال با NORM و رسم با PYPLOTT



تابع چگالی احتمال گوسی چند متغیره

$$p(\underline{x}) = \frac{1}{(2\pi)^{\frac{\ell}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} (\underline{x} - \underline{\mu})^T \Sigma^{-1} (\underline{x} - \underline{\mu})\right)$$

خط زیر به معنای بردار است.

Σ = covariance matrix

μ = mean vector

$|\cdot|$ = determinant

l = number of features (variables)

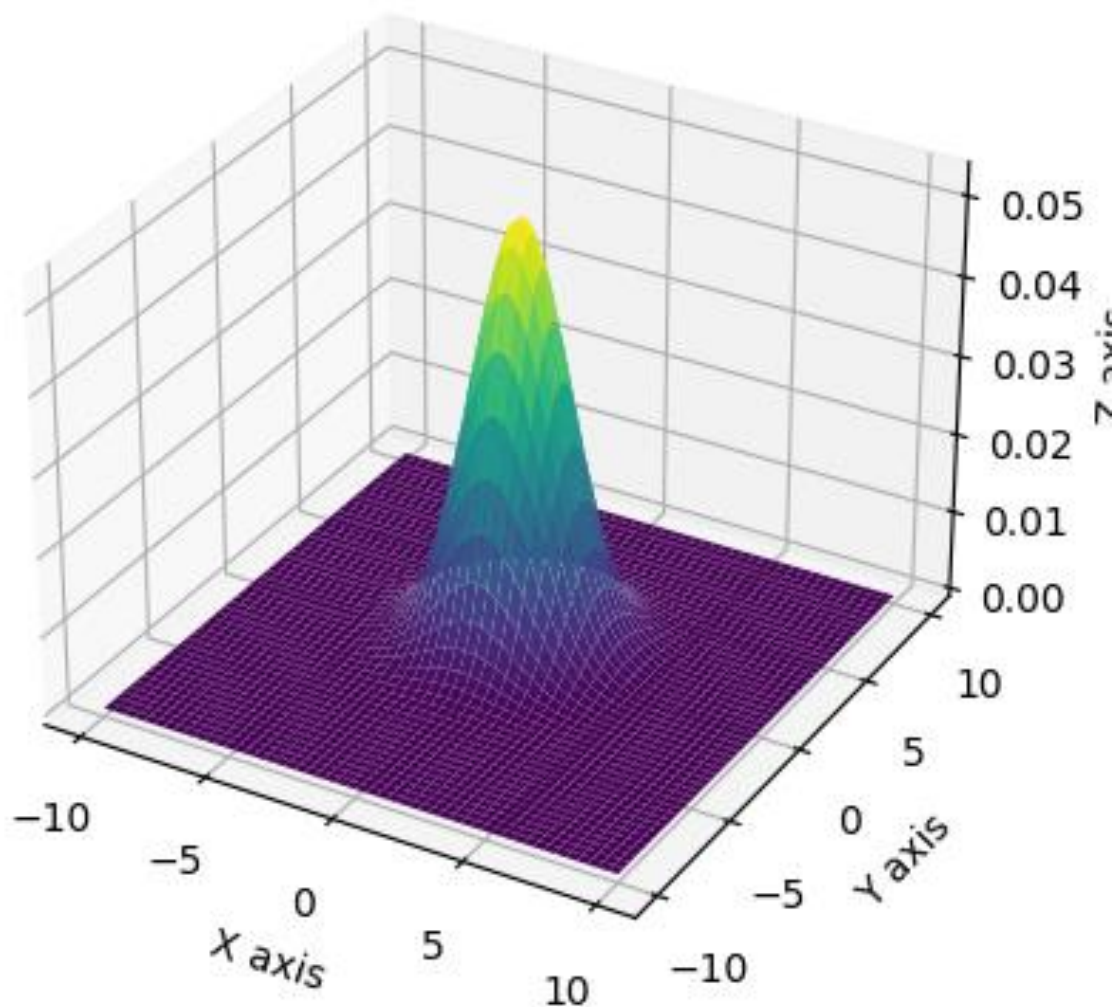
محاسبه تابع چگالی نرمال چند متغیره و رسم رویه

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import multivariate_normal
from mpl_toolkits.mplot3d import Axes3D
mu_x = 0 # mean of x # define parameters for x and y distributions
variance_x = 3 # variance of x
mu_y = 1 # mean of y
variance_y = 3 # variance of y
# define a grid for x and y values
x = np.linspace(-10, 10, 500) # generate 500 points between -10 and 10 for x
y = np.linspace(-10, 10, 500) # generate 500 points between -10 and 10 for y
X, Y = np.meshgrid(x, y) # create a grid for (x,y) pairs
# create an empty array of the same shape as X to hold the (x, y) coordinates
pos = np.empty(X.shape + (2,))
# fill the pos array with the x and y coordinates
pos[:, :, 0] = X
pos[:, :, 1] = Y
# create a multivariate normal distribution using the defined parameters
rv = multivariate_normal([mu_x, mu_y], [[variance_x, 0], [0, variance_y]])
# create a new figure for 3D plot
# ...
```

محاسبه تابع چگالی نرمال چند متغیره ...

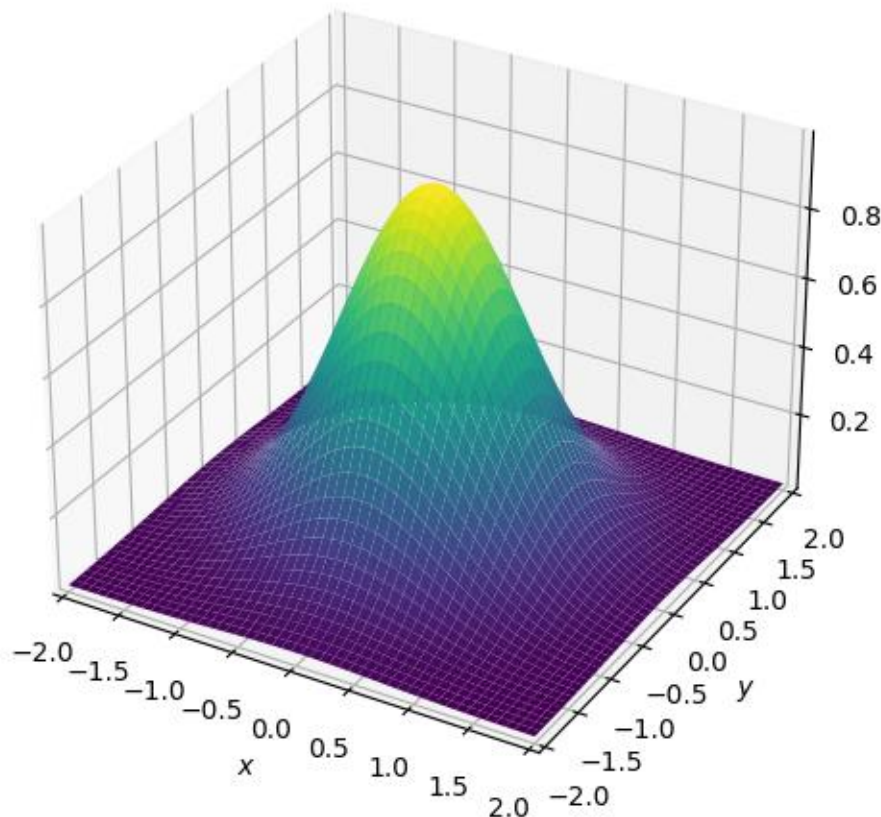
```
# ...  
# create a new figure for 3D plot  
fig = plt.figure()  
# add a 3D subplot to the figure  
ax = fig.add_subplot(projection='3d')  
# create a 3D surface plot of the multivariate normal distribution  
ax.plot_surface(X, Y, rv.pdf(pos), cmap='viridis', linewidth=0)  
# set labels for the axes  
ax.set_xlabel('X axis')  
ax.set_ylabel('Y axis')  
ax.set_zlabel('Z axis')  
# display the 3D plot  
plt.show()
```

محاسبه تابع چگالی نرمال چند متغیره ...



رسم تابع سه بعدی با استفاده از SYMPY (رسم با استفاده از معادله و سمبول)

```
from sympy.plotting import plot3d  
plot3d( $\exp(-(x^2+y^2))$ , ('x', -2,2), ('y', -2,2))
```



dition