Machine Learning Foundations

Chapter 6: Probability

Harish Guruprasad Ramaswamy
IIT Madras

Outline for Chapter 6 : Probability

6.1 : Discrete Random Variables

6.2 : Continuous Random Variables

6.3 : Maximum Likelihood and other advanced topics

Outline for Chapter 6 : Probability

6.1 : Discrete Random Variables

6.2 : Continuous Random Variables

6.3 : Advanced topics

1. Bivariate and Multivariate normal

2. Estimation of parameters using ML

3. Gaussian Mixture Models and Expectation Maximisation

4. Law of Large Numbers

Standard Normal Vector

$Z_1 \sim N(0,1), \quad Z_2 \sim N(0,1), \quad \ldots, \quad Z_d \sim N(0,1)$

$$Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_d \end{bmatrix}$$

$$f_Z(y) = \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} y_i^2\right)$$

$$f_Z(y) = \frac{1}{(2\pi)^{d/2}} \cdot \exp\left(-\frac{1}{2} \|y\|^2\right)$$

Simple Linear Transform of 2D-Normal

Let $\rho \in [-1,1]$. $X_1 = Z_1$; $X_2 = \rho Z_1 + \sqrt{1-\rho^2}\, Z_2$

$$X = \underbrace{\begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix}}_{A} Z \quad ; \quad Z = \underbrace{\begin{bmatrix} 1 & 0 \\ \frac{-\rho}{\sqrt{1-\rho^2}} & \frac{1}{\sqrt{1-\rho^2}} \end{bmatrix}}_{A^{-1}} X$$

$$\operatorname{Det}(A) = \sqrt{1-\rho^2} \quad ; \quad \operatorname{Det}(A^{-1}) = \frac{1}{\sqrt{1-\rho^2}}$$

$$E[X_1] = E[X_2] = 0$$

$$\operatorname{Cov}[X_1, X_2] = E[X_1 X_2] - E[X_1]E[X_2] = E[X_1 X_2]$$

$$= E\left[\rho Z_1^2 + \sqrt{1-\rho^2}\, Z_1 Z_2\right] = \rho\, E[Z_1^2] = \rho$$

Simple Linear Transform of 2D-Normal

$$Var[X_1] = 1$$

$$Var[X_2] = E\left[\left(\rho Z_1 + \sqrt{1-\rho^2}\, Z_2\right)^2\right] = \rho^2 + (1-\rho^2) \cdot 1$$

$$\therefore \; E[X] = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$cov[X] = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} = AA^T \quad ; \quad Det(\Sigma) = 1-\rho^2$$

$$\left[cov[X]\right]^{-1} = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} ; \; Det(\Sigma^{-1}) = \frac{1}{1-\rho^2}$$

Simple Linear Transform of 2D-Normal

$$f_X(x) = f_Z(A^{-1}x) \cdot |\mathrm{Det}(A^{-1})|$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}} \cdot \exp\left(-\frac{1}{2}\|A^{-1}x\|^2\right)$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2} x^T {A^{-1}}^T A^{-1} x\right)$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} [x_1 \quad x_2] \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right)$$

Simple Linear Transform of 2D-Normal

$$f_X(x) = \frac{1}{2\pi\sqrt{1-\rho^2}} \cdot \exp\left(-\frac{1}{2}\|A^{-1}x\|^2\right)$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2}\left(x_1^2 + \left(\frac{-\rho}{\sqrt{1-\rho^2}}x_1 + \frac{1}{\sqrt{1-\rho^2}}x_2\right)^2\right)\right]$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2}x_1^2\right) \cdot \exp\left(-\frac{1}{2(1-\rho^2)}(x_2-\rho x_1)^2\right)$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_1^2\right) \cdot \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left(\frac{-1}{2(1-\rho^2)}(x_2-\rho x_1)^2\right)$$

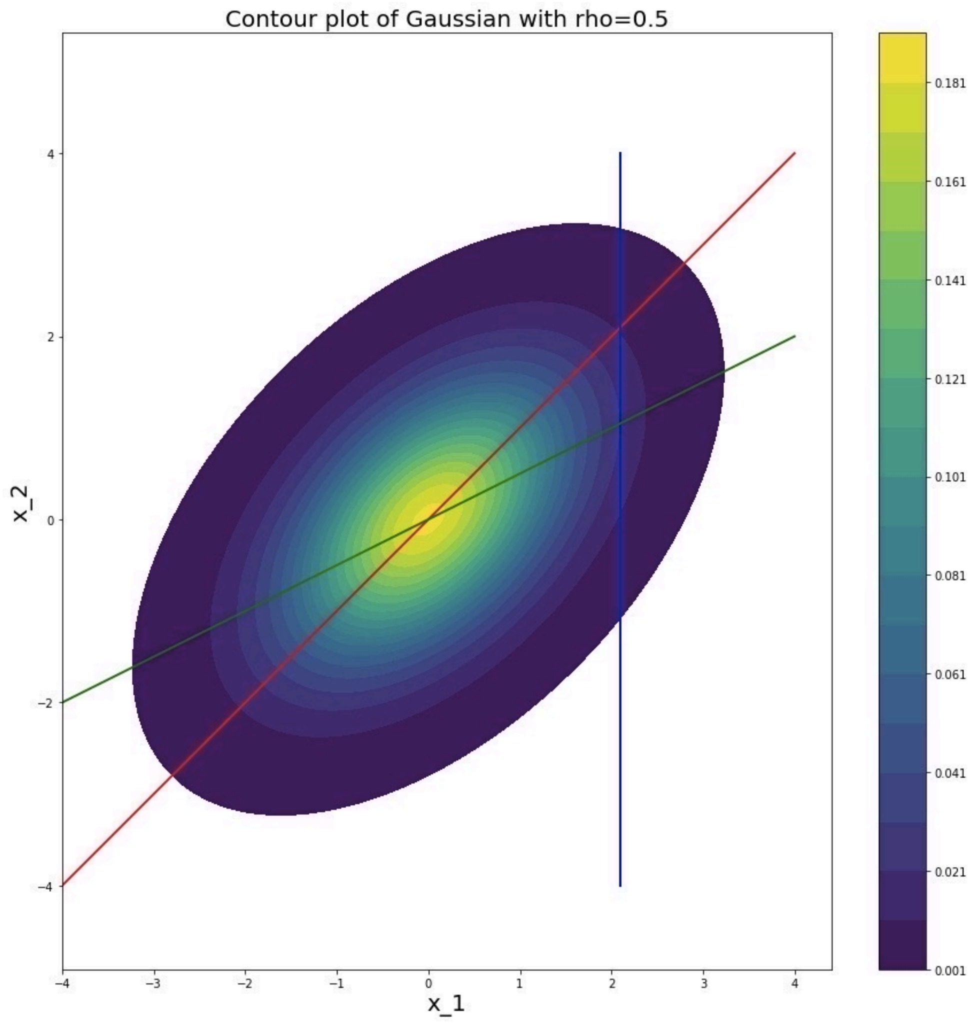
Simple Linear Transform of 2D-Normal

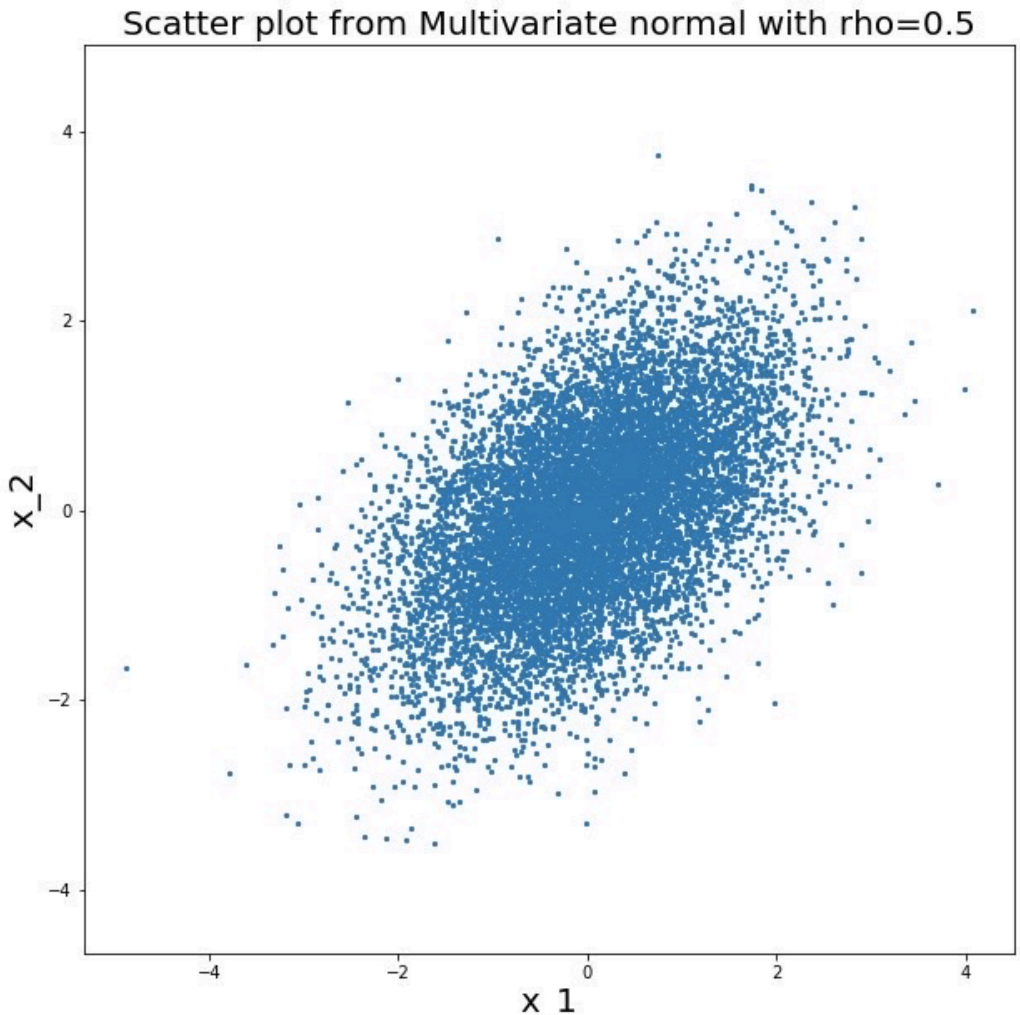
$$f_X(x) = \underbrace{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_1^2\right)}_{N(x_1 \mid 0, 1)} \cdot \underbrace{\frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} \exp\left(\frac{-1}{2(1-\rho^2)}(x_2 - \rho x_1)^2\right)}_{N(x_2 \mid \rho x_1,\, 1-\rho^2)}$$

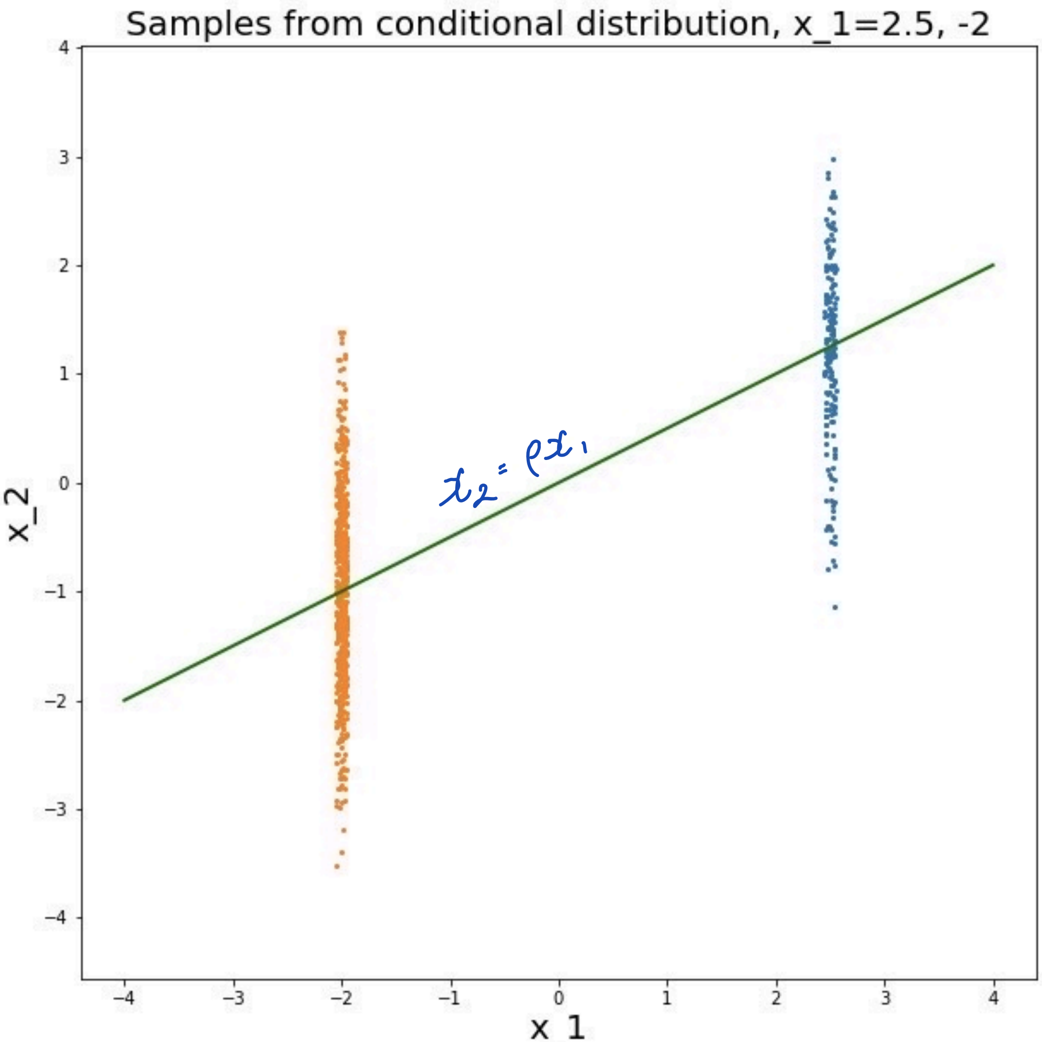
$$= f_{X_1}(x_1) \cdot f_{X_2 \mid X_1}(x_2 \mid x_1)$$

$$X_1 \sim N(0, 1)$$

$$X_2 \mid X_1 = x_1 \sim N(\rho x_1, 1-\rho^2)$$


Contour plot of Gaussian with rho=0.5
x_2
x_1
4
2
0
−2
−4
−4
−3
−2
−1
0
1
2
3
4
0.181
0.161
0.141
0.121
0.101
0.081
0.061
0.041
0.021
0.001


Scatter plot from Multivariate normal with rho=0.5
x_2
x_1
4
2
0
−2
−4
−4
−2
0
2
4



Samples from conditional distribution, x_1=2.5, -2
x_1
x_2
$x_2 = \rho x_1$

The Bivariate Normal

$$X = AZ$$

$$f_X(x) = f_Z(A^{-1}x) \cdot \left| \mathrm{Det}(A^{-1}) \right|$$

Let $\Sigma = AA^T = E[XX^T] = \mathrm{cov}[X]$

then $\mathrm{Det}\, A^{-1} = \dfrac{1}{\mathrm{Det}\, A} = \dfrac{1}{\sqrt{\mathrm{Det}(\Sigma)}}$

$$f_X(x) = \frac{1}{2\pi \sqrt{\mathrm{Det}(\Sigma)}} \exp\left(-\frac{1}{2} x^T \Sigma^{-1} x\right)$$

The Bivariate Normal

$$\Sigma = A A^T = \begin{bmatrix} a^2 & \rho ab \\ \rho ab & b^2 \end{bmatrix}$$

a, b +ve

$\rho \in [-1, 1]$

The Bivariate Normal

$$X = AZ + \mu$$

$$Z = A^{-1}(X - \mu)$$

$$\Sigma = AA^T$$

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

Multivariate Normal

$$X = AZ + \mu$$

$$Z = A^{-1}(X - \mu)$$

$$\Sigma = AA^T$$

$$f_X(x) = \frac{1}{(2\pi)^{d/2}\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

$$X \sim N(\mu, \Sigma)$$

Some Properties of the Multivariate Normal

Let $X \sim N(\mu, \Sigma)$

i) $Y = a^T X \sim N(a^T\mu, a^T \Sigma a)$

(ii) $Y = AX \sim N(A\mu, A \Sigma A^T)$

e.g.: $\Sigma = \begin{bmatrix} a^2 & \rho ab \\ \rho ab & b^2 \end{bmatrix}$; $\mu = 0$; $A = \begin{bmatrix} 1/a & 0 \\ 0 & 1/b \end{bmatrix}$

Then $Y = AX \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right)$

(iii) X_i, X_j are independent $\iff$ $\Sigma_{i,j} = 0$