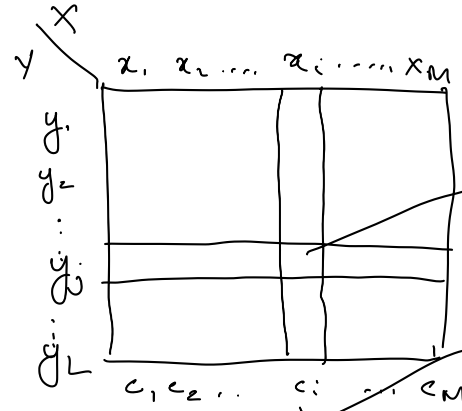
Probability Theory Background

Random Variables X & Y

$X \rightarrow x_1, x_2, \ldots x_M$ $\qquad$ $Y \rightarrow y_1, y_2, \ldots y_L$

Joint Distribution $p(X,Y)$



$p(X=x_i, Y=y_j) = p(x_i, y_j)$

$c_i = \sum_{j=1}^{L} n_{ij}$

N trials, let n_{ij} be number of times we observe $X=x_i$ & $Y=y_j$

As $N \to \infty$, $p(x_i, y_j) = \dfrac{n_{ij}}{N}$

marginal distribution

$$\boxed{p(X=x_i) = \sum_{j=1}^{L} p(X=x_i, Y=y_j) = \sum_{j=1}^{L} \frac{n_{ij}}{N} = \frac{c_i}{N}}$$

Sum Rule

Conditional Probability of $Y=y_j$ given that $X=x_i$

$$p(Y=y_j \mid X=x_i) = \boxed{\frac{n_{ij}}{c_i}}$$

$p(y_j \mid x_i)$

$$p(y_j, x_i) = \frac{n_{ij}}{N} = \boxed{\frac{n_{ij}}{c_i}} \cdot \frac{c_i}{N}$$

Product Rule

$$\boxed{p(Y=y_j, X=x_i) = p(Y=y_j \mid X=x_i)\, p(X=x_i)}$$

Sum Rule: $\boxed{p(X) = \sum_{Y} p(X,Y)}$

Product Rule $\boxed{p(X,Y) = p(Y|X)\, p(X)}$

$$p(y|x)\,p(x) = p(x,y) = p(x|y)\,p(y)$$

Bayes Rule : $p(y|x) = \dfrac{p(x|y)\,p(y)}{p(x)}$

evidence / "data likelihood"

prior

posterior

$$= \frac{p(x|y)\,p(y)}{\sum_y p(x,y)} = \frac{p(x|y)\,p(y)}{\sum_y p(x|y)\,p(y)}$$

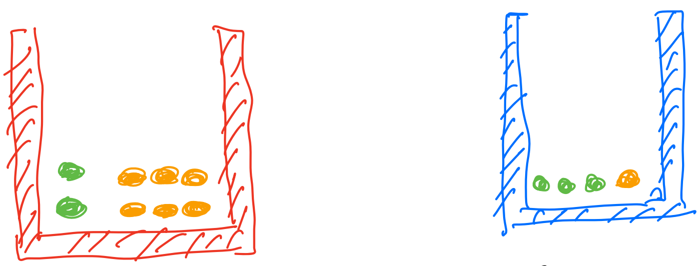
Independence : $p(x_i, y_j) = p(x_i)\,p(y_j)$, $p(y_j|x_i) = p(y_j)$

Example : Two boxes : blue & red

2 apples + 6 oranges

3 apples + 1 orange

Two fruit : Apples & Oranges



$p(r) = 0.4 = \frac{2}{5}$

$p(b) = 0.6 = \frac{3}{5}$

$p(a|r) = \frac{1}{4}$, $p(o|r) = \frac{3}{4}$

$p(a|b) = \frac{3}{4}$, $p(o|b) = \frac{1}{4}$

F \ B	r	b	
a	1/10	9/20	11/20
o	3/10	3/20	9/20
	2/5 = $p(r)$	12/20 = 3/5 = $p(b)$	

$\boxed{p(a)\,p(r|a)}$

$p(a,r) = p(a|r)\,p(r) = \frac{1}{4}\cdot\frac{2}{5} = \frac{1}{10}$

$p(o,r) = p(o|r)\,p(r) = \frac{3}{4}\cdot\frac{2}{5} = \frac{3}{10}$

$p(a,b) = p(a|b)\,p(b) = \frac{3}{4}\cdot\frac{3}{5} = \frac{9}{20}$

$p(o,b) = p(o|b)\,p(b) = \frac{1}{4}\cdot\frac{3}{5} = \frac{3}{20}$

$p(a) = p(a,r) + p(a,b) = \frac{1}{10} + \frac{9}{20} = \frac{11}{20}$; $1 - \frac{11}{20} = \frac{9}{20}$

$p(o) = p(o,r) + p(o,b) = \frac{3}{10} + \frac{3}{20} = \frac{9}{20}$

$$p(r|o) = \frac{p(o|r)\,p(r)}{p(o)} = \frac{3/4 \cdot 2/5}{9/20} = \frac{3/10}{9/20} = \frac{3}{10}\cdot\frac{20}{9} = \frac{2}{3}$$

Probabilities wrt continuous variables, $x \in \mathbb{R}$ $(x \in \mathbb{R}^d)$

pdf = probability density function $p(x)$

$p(x) \geq 0$, $\int_{-\infty}^{\infty} p(x)\,dx = 1$

Sum Rule : $p(x) = \int_{-\infty}^{\infty} p(x,y)\,dy$

Product Rule : $p(x,y) = p(y|x)p(x) = p(x|y)\,p(y)$

Expectation (Mean)

$E[f(x)] = \int p(x)\,f(x)\,dx \quad \left(= \sum_x p(x)\,f(x) \text{ in discrete case} \right)$

$E[x] = \int x\,p(x)\,dx \longrightarrow E[x+y] = E[x] + E[y]$

<u>Variance</u>

$$\begin{aligned}
Var[f(x)] &= E\left[(f(x) - E[f(x)])^2 \right] \\
&= E\left[(f(x))^2 + (E[f(x)])^2 - 2f(x)E[f(x)] \right] \\
&= E[(f(x))^2] + E\left[E[f(x)]\right]^2 - 2E[f(x)E[f(x)]] \\
&= E[(f(x))^2] + \underbrace{E[(f(x))]^2 - 2E[f(x)]^2} \\
&= E[(f(x))^2] - E[(f(x))]^2
\end{aligned}$$

$Var(x) = E[(x - E[x])^2] = E[x^2] - (E[x])^2$

covariance $\equiv$ cov

$$\begin{aligned}
cov(x,y) &= E\left[\{x - E[x]\} \{y - E[y]\} \right] \\
&= E[xy] - E[x]E[y]
\end{aligned}$$

What if x and y are independent? $p(x,y) = p(x)p(y)$

$cov(x,y) = E[xy] - E[x]E[y]$

$\int \underbrace{p(x,y)}_{p(x)p(y)} xy\,.dy\,dx$

$\int xp(x)\,dx\,p(y$

$E[x].E[y]$

$cov(x,y) = E[x]E[y] - E[x]E[y] = 0$ if x & y are independent

$cov(x,y) = 0$ if x & y are independent

Gaussian Distribution / Normal Distribution

$x \in \mathbb{R}$

$$p(x \mid \mu, \sigma^2) = p(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}(x-\mu)^2/\sigma^2}$$

μ = mean or expectation

σ^2 = variance

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

$$E[(x-\mu)^2] = E[x^2] - \mu^2 = \sigma^2$$

$\mu^2 = (E[x])^2$

$$x \in \mathbb{R}^d, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}, \quad E[x] = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_d \end{bmatrix} = \mu$$

$\mu \in \mathbb{R}^d$

$$E[(x_i-\mu_i)^2] = \sigma_i^2$$

$$p(x) = p(x_1, x_2, \dots x_d)$$

Suppose x_i is independent of x_j $\forall\, i \neq j$

$$p(x) = p(x_1)p(x_2)\cdots p(x_d) = \prod_{i=1}^{d} p(x_i)$$

$$= \frac{1}{(\sqrt{2\pi})^d \sigma_1\sigma_2\cdots\sigma_d} \prod_{i=1}^{d} e^{-\frac{1}{2}(x_i-\mu_i)^2/\sigma_i^2}$$

$e^a e^b = e^{a+b}$

$$= \frac{1}{(2\pi)^{d/2}\sigma_1\sigma_2\cdots\sigma_d} \cdot e^{-\frac{1}{2}\sum_{i=1}^{d}(x_i-\mu_i)^2/\sigma_i^2}$$

$$\Sigma = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \sigma_2^2 & \\ 0 & & \ddots\ \sigma_d^2 \end{bmatrix}, \quad \det(\Sigma) = \sigma_1^2\sigma_2^2\cdots\sigma_d^2 = (\sigma_1\sigma_2\cdots\sigma_d)^2$$

$$x-\mu = \begin{bmatrix} x_1-\mu_1 \\ x_2-\mu_2 \\ \vdots \\ x_d-\mu_d \end{bmatrix}$$

$$= \frac{1}{(2\pi)^{d/2}(\det(\Sigma))^{1/2}} \cdot$$

$$(x-\mu)^T(x-\mu) = \sum_{i=1}^{d}(x_i-\mu_i)^2$$

$$(x-\mu)^T\Sigma^{-1}(x-\mu) = \sum_{i=1}^{d}(x_i-\mu_i)^2/\sigma^2$$

$$, \quad \Sigma = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \sigma_2^2 & \\ 0 & & \ddots\ \sigma_d^2 \end{bmatrix}$$

Σ is diagonal

$$\boxed{p(x) = \frac{1}{(2\pi)^{d/2}(\det(\Sigma))^{1/2}} \cdot e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}}$$

General Case

$x \in \mathbb{R}^d$

Multivariate Gaussian Distribution

$$\boxed{p(x) = \frac{1}{(2\pi)^{d/2}(\det(\Sigma))^{1/2}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}}$$

where μ is the mean, ie. $\mu = E[x] = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_d \end{bmatrix}$, $\mu_i = E[x_i]$

Σ is the Covariance Matrix $= E[(x-\mu)(x-\mu)^T]$

Σ_{ij} is the covariance between x_i and x_j

Σ is $d \times d$, symmetric, positive definite

$\Sigma = V \Lambda V^T$ (eigenvalue decomposition/SVD)

(Λ is diagonal, $\Lambda_{ii} > 0$
$V^T V = V V^T = I$)

$\Sigma^{-1} = V \Lambda^{-1} V^T$

$\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu) = \frac{1}{2}(x-\mu)^T V \Lambda^{-1} \underbrace{V^T(x-\mu)}_{=z}$

$\boxed{z = V^T(x-\mu)}$

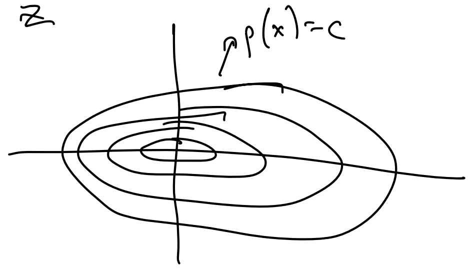
$\Rightarrow \frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu) = \frac{1}{2} z^T \Lambda^{-1} z$

$\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{bmatrix}$

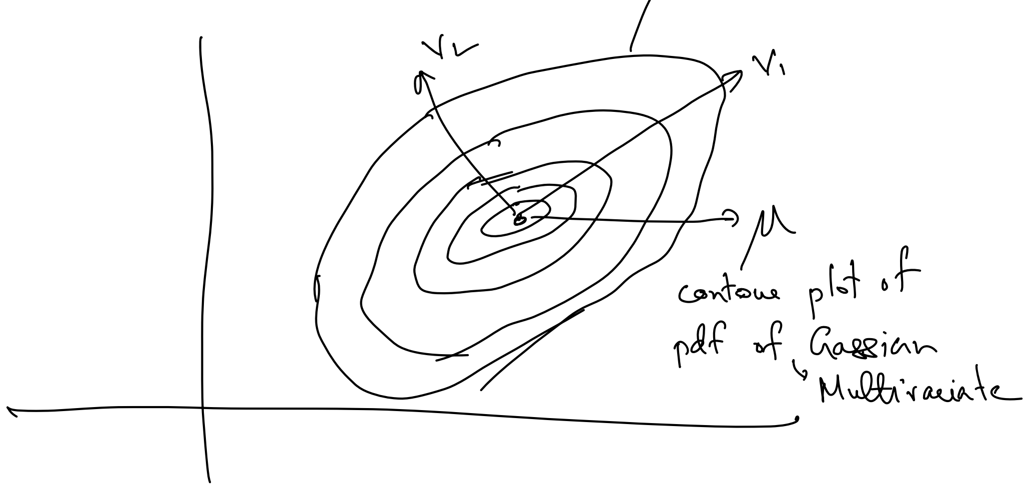
$p(x) = c \Rightarrow \frac{1}{2} z^T \Lambda^{-1} z = c'$

$\boxed{\frac{1}{2}\sum \frac{z_i^2}{\lambda_i} = c'}$ → Equation of an ellipse

z



$p(x) = c$


v_2
v_1
μ
contour plot of
pdf of Gassian
Multivariate