Lecture 2

Unsupervised Learning (Only data, no labels)

$I \in R^{28 \times 28} \rightarrow R^{784}$

1 2 3 4
5 6 7 8

28 $\boxed{2}$ 28 $\boxed{3}$

[0, 1 ... 255] grayscale range

↓

[0, 1]

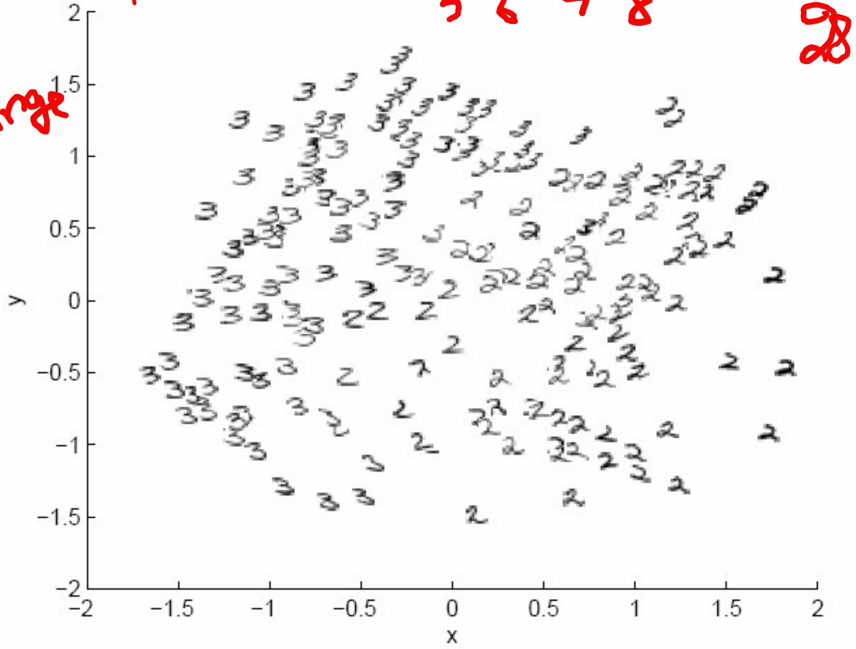
4K x 2K

10^6 dim

PCA

PCA: $I \in R^d \rightarrow R^p$

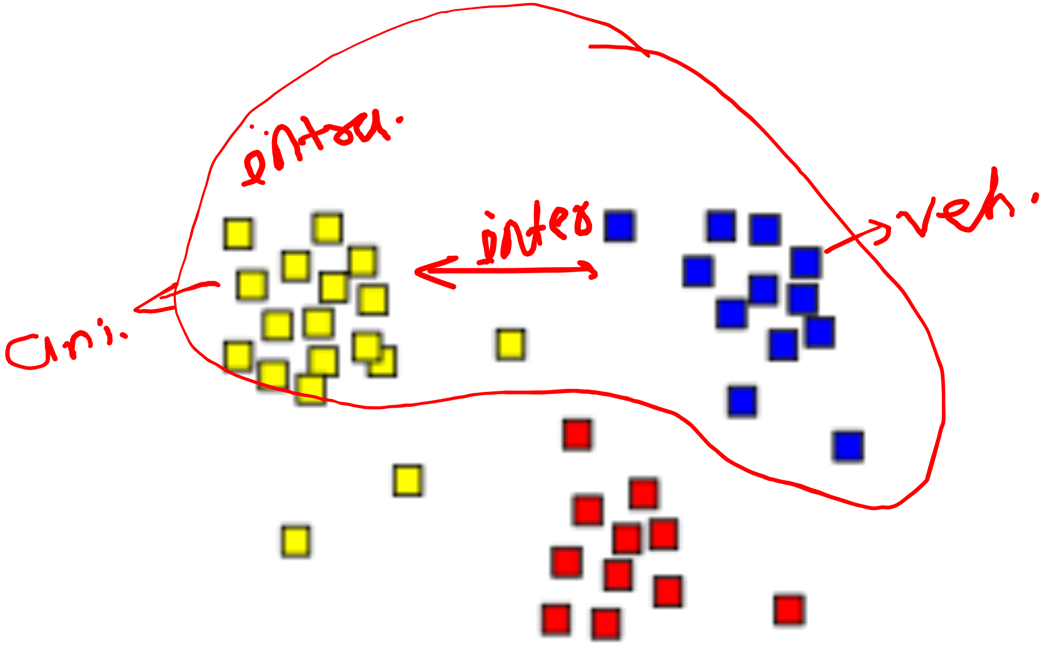
$p \ll d$



A canonical dimensionality reduction problem from visual perception. The input consists of a sequence of 64-dimensional vectors, representing the brightness values of 8 pixel by 8 pixel images of digits 2 and 3. Applied to $n = 400$ raw images. A two-dimensional projection is shown, with the original input images.

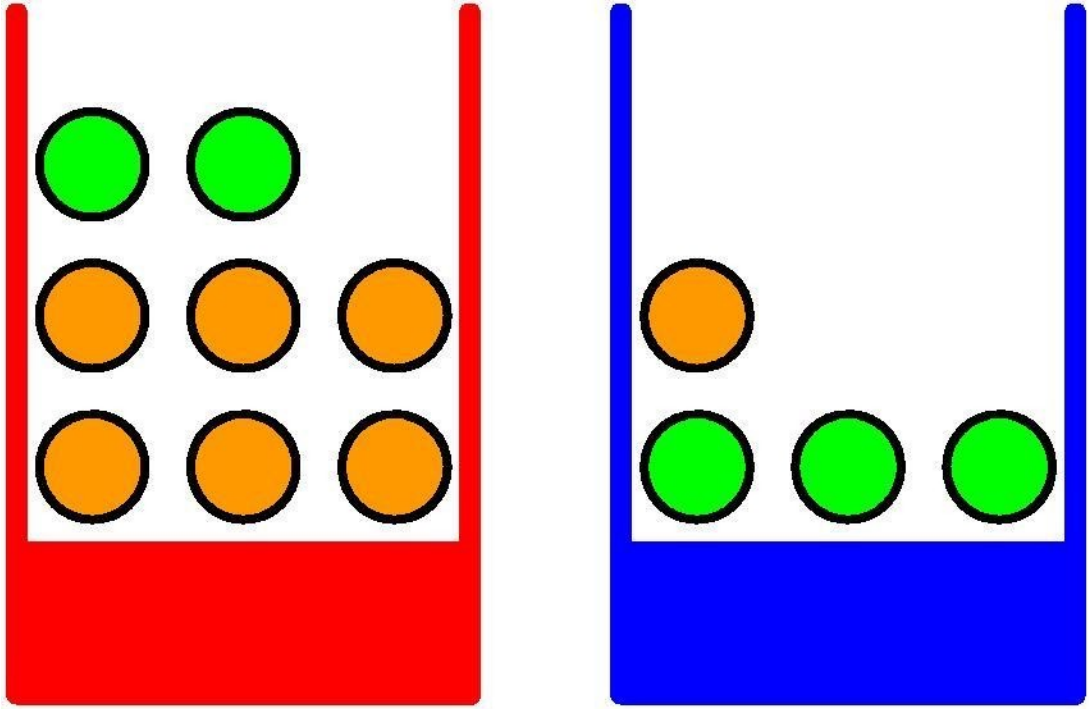
Clustering

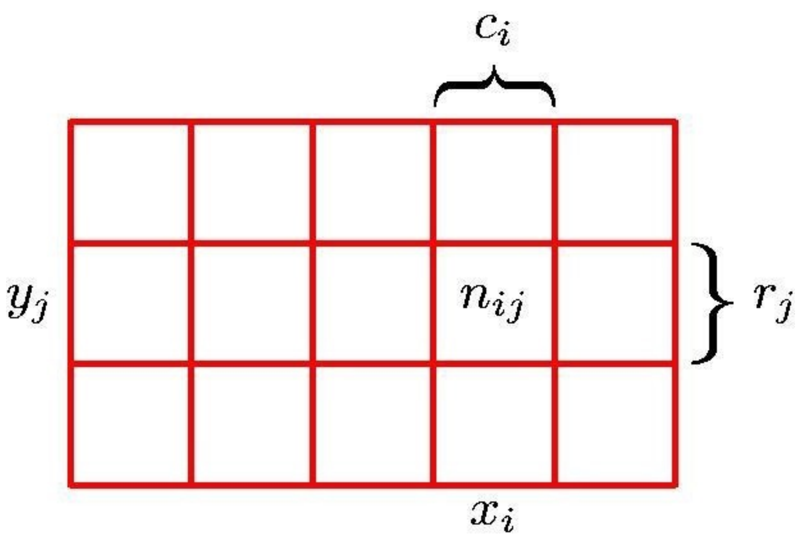
Organizing data into clusters such that there is
- high intra-cluster similarity
- low inter-cluster similarity
- Informally, finding natural groupings among objects.



Probability Theory

Apples and Oranges



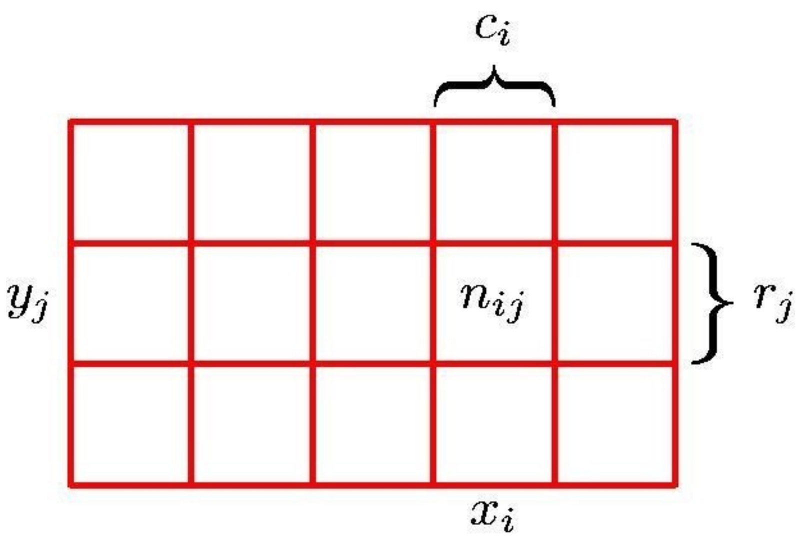


Let X and Y be random variables.

Let their be N trial during which we sample both of variables X and Y.

Let the number of times X=xi and Y=yj occur is nij.

Probability Theory



Marginal Probability

$$p(X = x_i) = \frac{c_i}{N}.$$

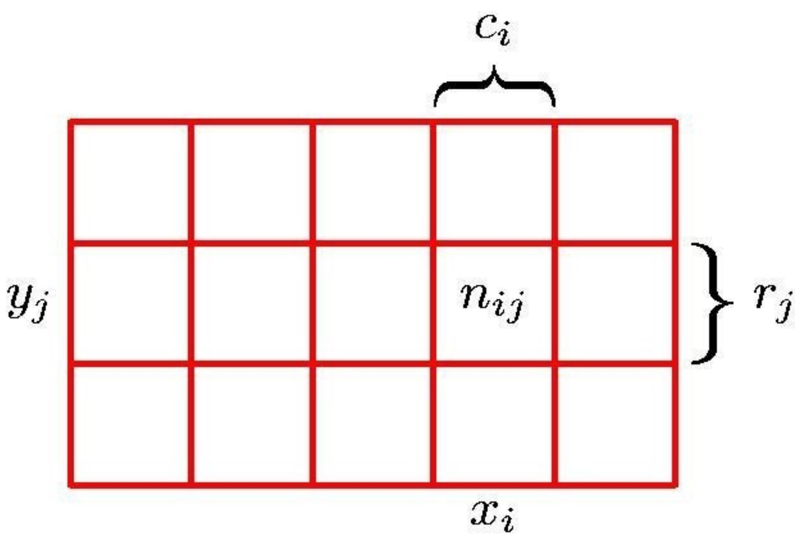
Joint Probability

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

Conditional Probability

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

Probability Theory



Sum Rule

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

Product Rule

$$
\begin{aligned}
p(X = x_i, Y = y_j) &= \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \cdot \frac{c_i}{N} \\
&= p(Y = y_j | X = x_i) p(X = x_i)
\end{aligned}
$$

The Rules of Probability

$$P(X) = \sum_Y P(X|Y)\,P(Y)$$

Sum Rule $$p(X) = \sum_Y p(X, Y)$$

Product Rule $$p(X, Y) = p(Y|X)p(X)$$

Bayes' Theorem

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)}$$

$$p(X) = \sum_{Y} p(X|Y)p(Y)$$

Ex. Consider red and blue boxes

B → r.v. related Box
F → r.v. " to fruit.

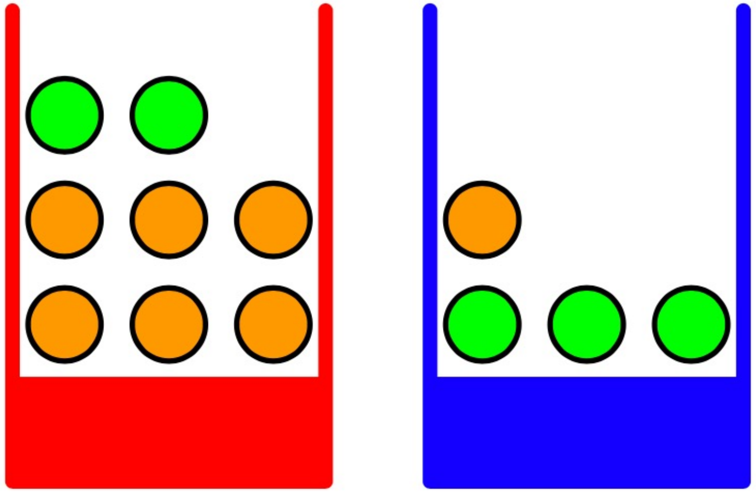
$p(B = r) = 4/10$ and $p(B = b) = 6/10$

$p(F = a | B = r) = 1/4$

$p(F = o | B = r) = 3/4$

$p(F = a | B = b) = 3/4$

$p(F = o | B = b) = 1/4.$



$P(B=r) = 4/10$

Contd.

$P(B=r|F=o) + P(B=b|F=o) = 1$

We are told that a piece of fruit has been selected and it is an orange, and we would like to know which box it came from.
Bayes: p(B=r|F=o) = p(F = o|B = r) p(B=r) / p(F=o)

p(F =o) = p(F =o|B=r)p(B=r)+p(F =o|B=b)p(B=b)

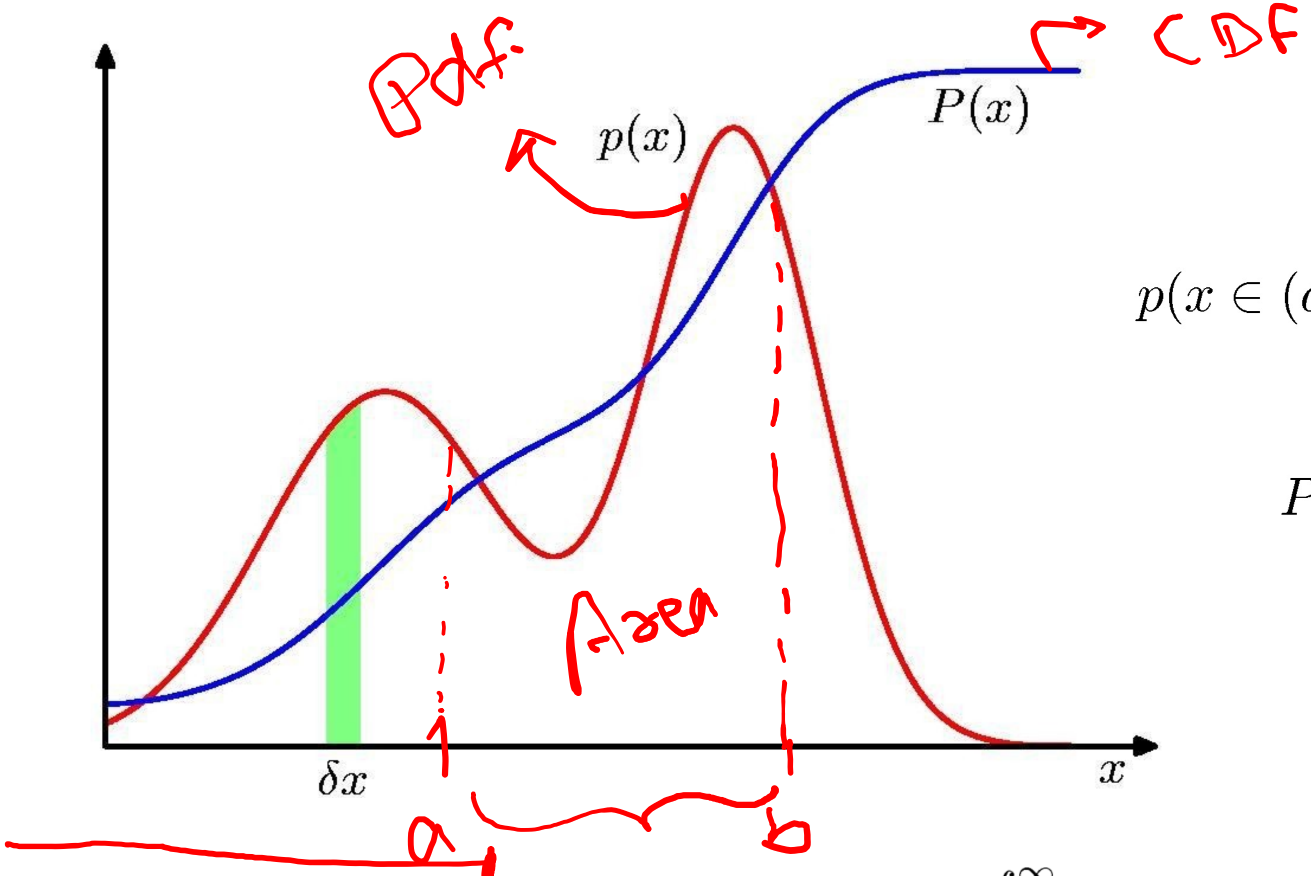
$$P(B=r|F=o) = \frac{P(F=o|B=r)\,P(B=r)}{P(F=o)}$$

$$= \frac{\frac{3}{4}\cdot\frac{4}{10}}{P(F=o|B=r)P(B=r) + P(F=o|B=b)P(B=b)}$$

$$P(B=b|F=o)$$

$$\frac{3/10}{\frac{3}{4}\cdot\frac{4}{10}+\frac{1}{4}\cdot\frac{6}{10}}$$

Probability Densities



$$p(x \in (a,b)) = \int_a^b p(x)\,\mathrm{d}x$$

$$P(z) = \int_{-\infty}^{z} p(x)\,\mathrm{d}x$$

Z $p(x) \geqslant 0$

$$\int_{-\infty}^{\infty} p(x)\,\mathrm{d}x = 1$$

Expectations

x is discrete

$$\mathbb{E}[f] = \sum_x p(x) f(x)$$

x is cont.

$$\mathbb{E}[f] = \underbrace{\int p(x) f(x) \, \mathrm{d}x}$$

f(x) function.

f(x) = x

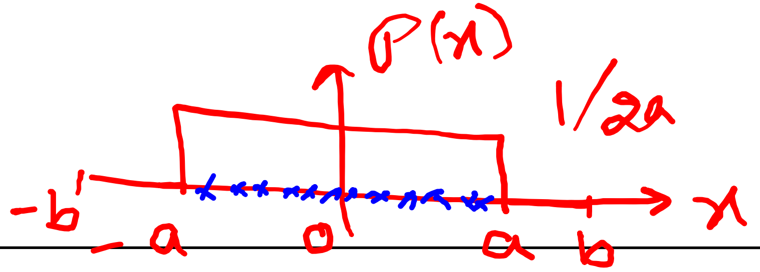
$$\mathbb{E}_x[f|y] = \sum_x p(x|y) f(x)$$

Conditional Expectation (discrete)

$$\mathbb{E}[f] \simeq \frac{1}{N} \sum_{n=1}^{N} f(x_n)$$

Approximate Expectation (discrete and continuous)

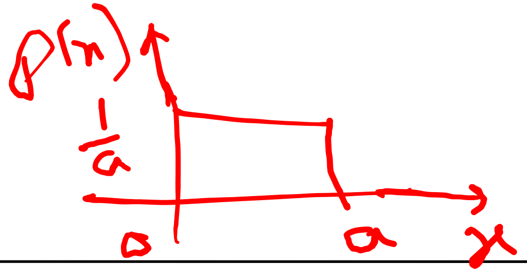
Ex. A uniform pdf in (-a,a).



$$E(x) = \int_{-a}^{a} x\, p(x)\, dx = \frac{1}{2a}\int_{-a}^{a} x\, dx = 0$$

$E(x) \rightarrow$ mean of ~~Pf~~ the random variable

$$E(x^2) = \int_{-a}^{a} p(x)\, x^2\, dx = \frac{1}{2a}\int_{-a}^{a} x^2\, dx$$



$x \in (-\infty, \infty)$

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

$$\mathrm{Var}(x) = E\left[(x-\mu)^2\right] = \int_{-\infty}^{\infty} (x-\mu)^2 \, p(x)\, dx$$

$$= \int_{-\infty}^{\infty} x^2 \, p(x)\, dx = E(x^2)$$

$\mu \to$ mean for x or $E(x)$

$\mu = E(x)$

$$E(x) \approx \frac{1}{N}\sum_{n=1}^{N} x_n$$ there are total N points given

p(Animal) Animal

$\in \mathbb{R}^{d\times 1}$

N samples.

x_n

p(Human) Human.

$\in \mathbb{R}^{d\times 1}$

N samples.

x_n

$$E(x-\mu)^2 \approx \frac{1}{N}\sum_{n=1}^{N} (x_n-\mu)^2$$

Variances and Covariances

$$\mathrm{var}[f] = \mathbb{E}\left[(f(x) - \mathbb{E}[f(x)])^2\right] = \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2$$

$$\mathrm{Var}(x) = E(x-\mu)^2 = E(x^2) - \mu^2$$

$$\begin{aligned} \mathrm{cov}[x, y] &= \mathbb{E}_{x,y}\left[\{x - \mathbb{E}[x]\}\{y - \mathbb{E}[y]\}\right] = \iint (x-\mu_x)(y-\mu_y)\, p(x,y)\, dx\, dy \\ &= \mathbb{E}_{x,y}[xy] - \mathbb{E}[x]\mathbb{E}[y] \end{aligned}$$

$$\begin{aligned} \mathrm{cov}[\mathbf{x}, \mathbf{y}] &= \mathbb{E}_{\mathbf{x},\mathbf{y}}\left[\{\mathbf{x} - \mathbb{E}[\mathbf{x}]\}\{\mathbf{y}^{\mathrm{T}} - \mathbb{E}[\mathbf{y}^{\mathrm{T}}]\}\right] \\ &= \mathbb{E}_{\mathbf{x},\mathbf{y}}[\mathbf{x}\mathbf{y}^{\mathrm{T}}] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{y}^{\mathrm{T}}] \end{aligned}$$

$$\iint \left[xy\,p(x,y) - \underbrace{y\mu_x p(x,y)}_{-\mu_x\mu_y} - x\mu_y p(x,y) + \mu_x\mu_y \right] dx\,dy$$

$$-\mu_x\mu_y - \mu_x\mu_y + \mu_x\mu_y$$

$$\iint \mu_x\, y\, \underbrace{p(x,y)}\, dx\, dy$$

$$= \mu_x \iint y\, \underbrace{p(y|x)\, p(x)}\, dx\, dy$$

$$= \mu_x \int y \underbrace{\int p(y|x)\, p(x)\, dx}_{p(y)}\, dy$$

$$= \mu_x \underbrace{\int y\, p(y)\, dy}_{\mu_y} = \mu_x \mu_y$$

Is cov. matrix symmetric?

Is it PSD?

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

$x = \{1, 2, 3\} \qquad y = \{-1, 3, 0\}$

$$\mathrm{Cov}(x,y) = E_{x,y}(xy) - \mu_x \mu_y$$

$$\approx \frac{1}{N}\sum_{n=1}^{N} x_n y_n - \frac{1}{N}\sum_{n=1}^{N} x_n \, \frac{1}{N}\sum_{n=1}^{N} y_n$$

$N = 3 \qquad \frac{5}{3} - 2 \cdot \frac{2}{3}$

$Y = \begin{bmatrix} \text{age} \\ \text{ht} \\ \text{wt.} \end{bmatrix}$ $\quad Y \leftarrow X$

$Cov(X, Y) \quad X \in \mathbb{R}^d \quad Y \in \mathbb{R}^d$

$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ $\quad x_1 \to$ hrs. of effort.
$x_2 \to$ Peer learning
$x_3 \to$ Stock Price

$X_1 \; X_2 \; \ldots \; X_{120}$

$\bar{X} \to$

3×120

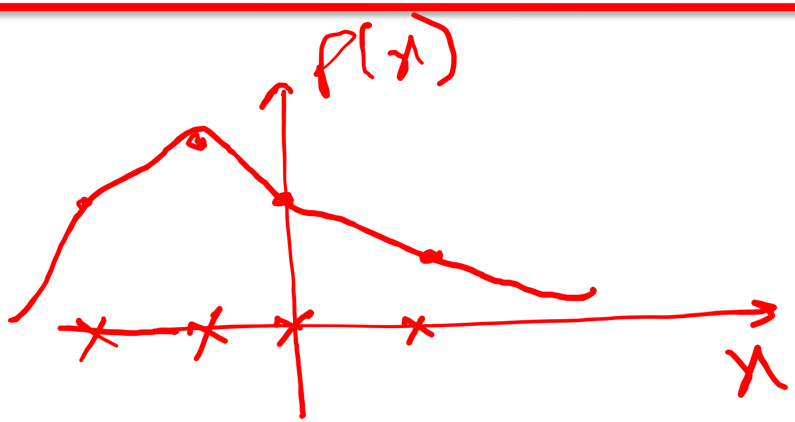
$E(X) \in \mathbb{R}^d = \begin{matrix} E(x_1) \\ E(x_2) \\ E(x_3) \end{matrix}$

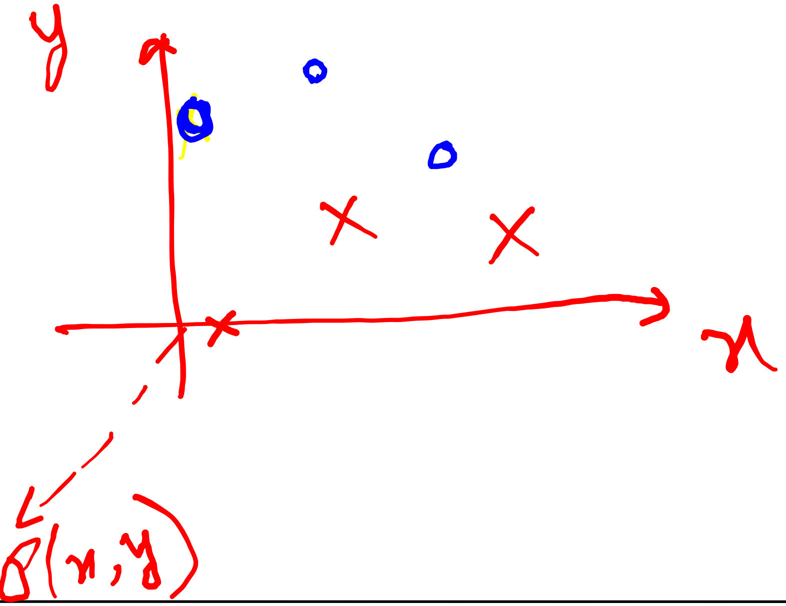
$Cov(X, Y)$

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

$\begin{bmatrix} \sigma^2_{x_1} & \sigma^2_{x_1,x_2} & \sigma^2_{x_1,x_3} \\ \sigma^2_{x_1,x_2} & \sigma^2_{x_2} & \sigma^2_{x_2,x_3} \\ \sigma^2_{x_1,x_2} & \downarrow & \sigma^2_{x_3} \end{bmatrix}$

$\sigma^2_{x_3,x_2}$


p(x)
x


y
x
p(x,y)