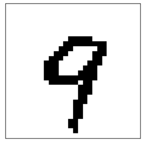
Covariance, Random Vectors

Contents

- Joint cdfs and pdfs
- Expectation of multiple RVs
- Covariance, correlation
- Random vectors
- Covariance matrix

Big Picture

- machine learning is about **learning functions from data**
- the inputs to the functions are called **features**

$$\boldsymbol{x} =$$ 

- the true output is called a label

$$y \in \{0, 1, 2, \ldots, 9\}$$

- labeled examples of features make up *training data*

$$\mathcal{D} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{n}$$

- useful characterization:

$$(\boldsymbol{x}_i, y_i) \stackrel{i.i.d.}{\sim} p(\boldsymbol{x}, y)$$

- if we have a good approximation of $p(\boldsymbol{x}, y)$, we can design a good function for classification:

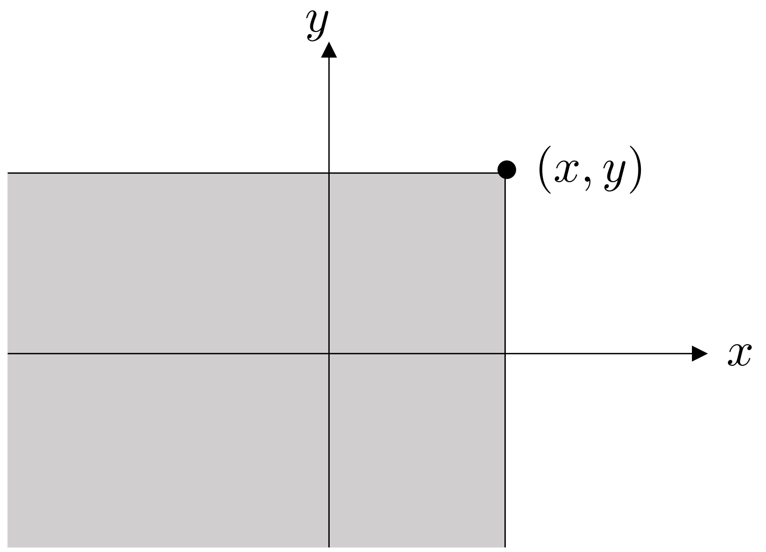
$$\widehat{y} = \arg\max_{y} p(y|\boldsymbol{x})$$

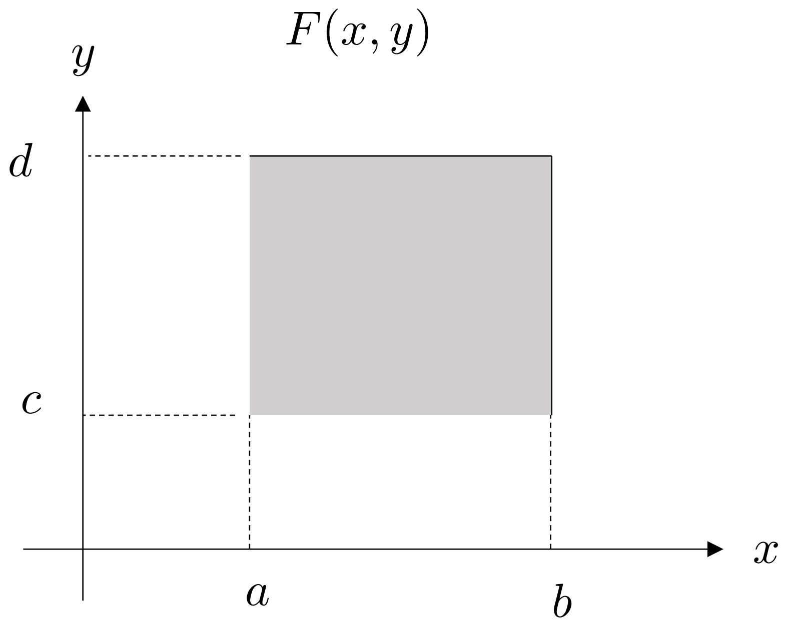
Joint cdfs

- let X and Y be two random variables
- X and Y are completely described by their *joint* cdf

$$F(x,y) = \mathbb{P}(X \leq x, Y \leq y) \text{ for all } x \in \mathbb{R}, y \in \mathbb{R}$$

- properties:
  - $F(x,y) \geq 0$
  - nondecreasing in both x, y
  - $\lim_{x\to\infty} F(x,y) = F(y)$
  - independent $\Leftrightarrow F(x,y) = F(x)F(y)$





- example $\mathbb{P}(a < X \leq b, c < Y \leq d)$

$$= F(b,d) - F(a,d) - F(b,c) + F(a,c)$$

Joint pdfs

- X and Y are completely described by their *joint* cdf

$$F(x, y) = \mathbb{P}(X \leq x, Y \leq y) \text{ for all } x \in \mathbb{R}, y \in \mathbb{R}$$

- we can define the joint pdf (if it exists) as a function $f(x, y)$ that satisfies:

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) du dv \text{ for all } x, y \in \mathbb{R}$$

- if X and Y are ***jointly* continuous**, then they are completely described by their *joint* pdf

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} \qquad f(x, y) = \lim_{\Delta x \Delta y \to 0} \frac{\mathbb{P}(x < X \leq x + \Delta x, y < Y \leq y + \Delta y)}{\Delta x \Delta y}$$

Properties of joint pdfs

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}$$

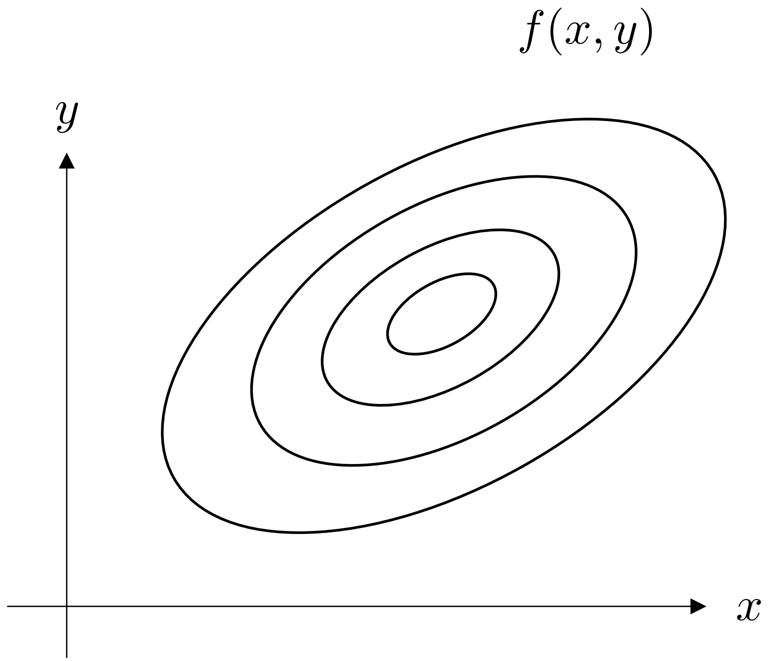
- properties:
  - $f(x,y) \geq 0$
  - $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$
  - $\mathbb{P}((X,Y) \in A) = \int_{(x,y)\in A} f(x,y)dxdy$

- the *marginal pdf* of X can be computed as:

$$f(x) = \int_{-\infty}^{\infty} f(x,y)dy$$

- X, Y are independent if and only if

$$f(x,y) = f(x)f(y) \text{ for all } x,y$$

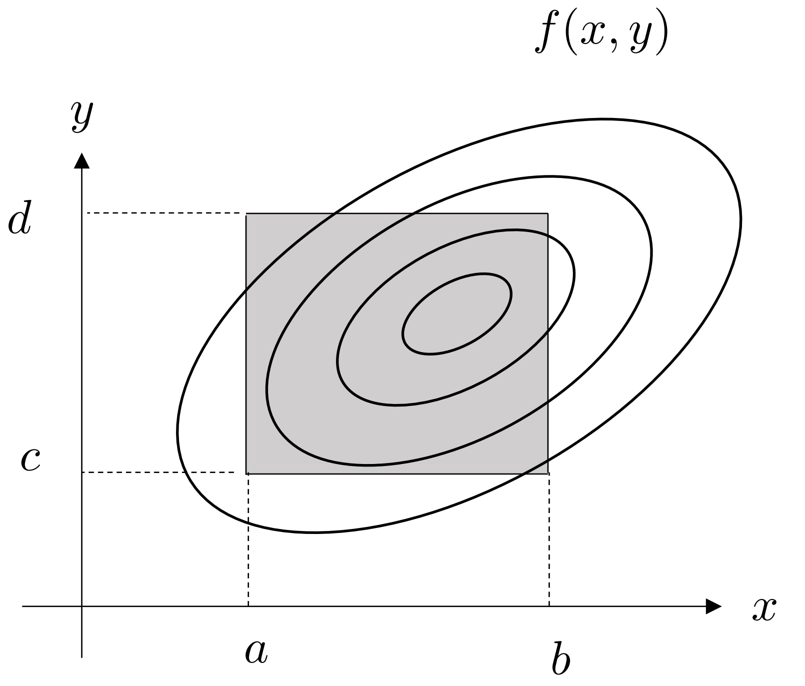


Example

$$\mathbb{P}((X,Y) \in A) = \int_{(x,y)\in A} f(x,y)dxdy$$

$$A = \{(x,y) : a \leq x \leq b, c \leq y \leq d\}$$

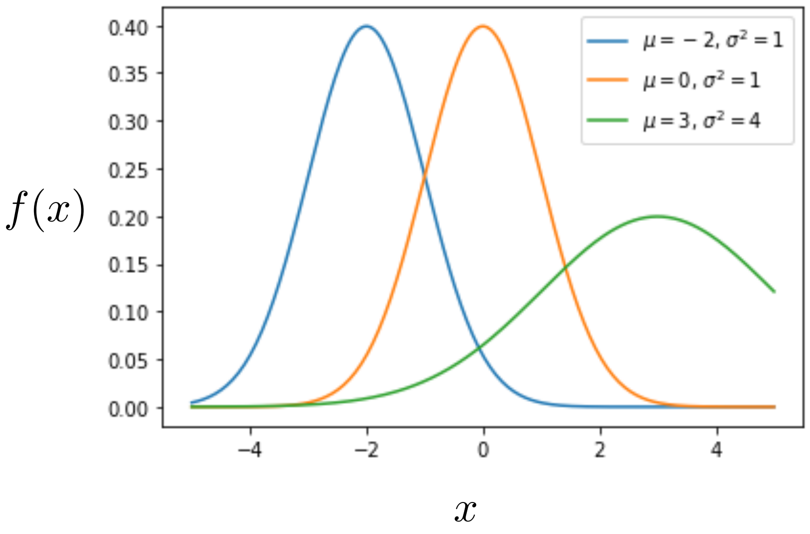
$$\mathbb{P}((X,Y) \in A) = \int_a^b \int_c^d f(x,y)dxdy$$



Example – bivariate Normal

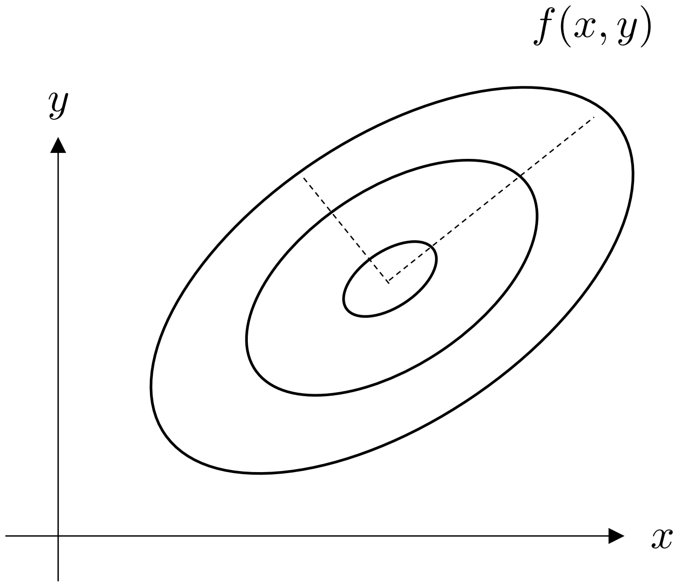
- Gaussian: $X \sim \mathcal{N}(\mu, \sigma^2)$

$$f(x) = \frac{\exp\left[\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]}{\sqrt{2\pi}\,\sigma}$$



- Bivariate (jointly) Gaussian:

$$f(x,y) = \frac{\exp\left[\frac{-1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\}\right]}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}$$



$\mu_x := E[X] \qquad \sigma_x^2 := \mathrm{var}(X)$

$\mu_y := E[Y] \qquad \sigma_y^2 := \mathrm{var}(Y)$

$\rho :=$ correlation

Expectation, Covariance, Correlation

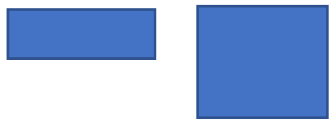
Expectation

- expectation: let $X, Y \sim p(x, y)$ and consider a function $g(x, y)$:

$$E[g(X,Y)] = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} g(x,y)p(x,y)$$

$$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy$$

- example: average area of a random rectangle



$$\text{area} = xy$$

$$g(x,y) = xy$$

	x			
		0.5	1	2
	1	$\frac{1}{8}$	$\frac{1}{8}$	0
y	2	0	$\frac{1}{2}$	0
	3	0	0	$\frac{1}{4}$

Correlation and Covariance

- the *variance* of a random variable X is defined as

$$\mathrm{var}(X) = E[(X - E[X])^2]$$

- *variance* of a sum of RVs:

$$\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y) + 2\mathrm{cov}(X,Y)$$

- *covariance* of two RVs:

$$\begin{aligned}\mathrm{cov}(X,Y) &= E\left[(X - E[X])(Y - E[Y])\right] \\ &= E[XY] - E[X]E[Y]\end{aligned}$$

- *correlation* of two RVs:

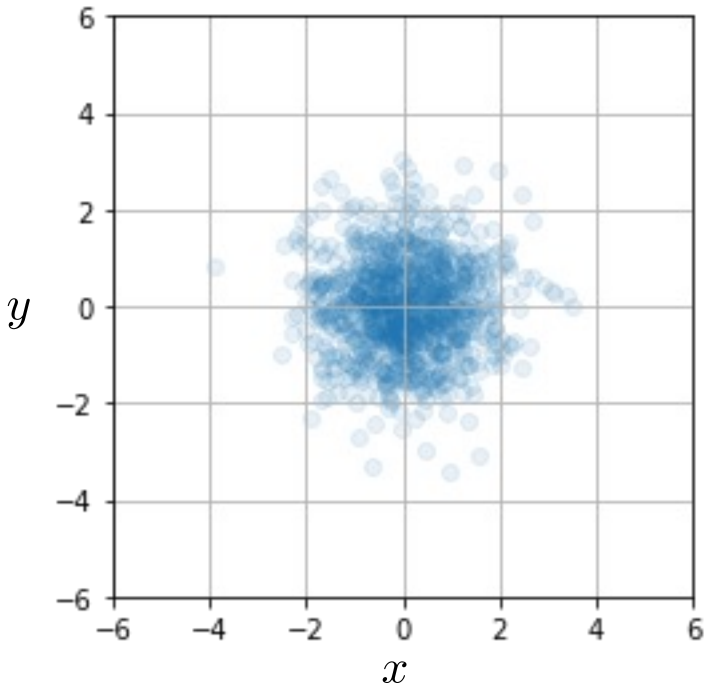
$$E[XY] = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} xyp(x,y)$$

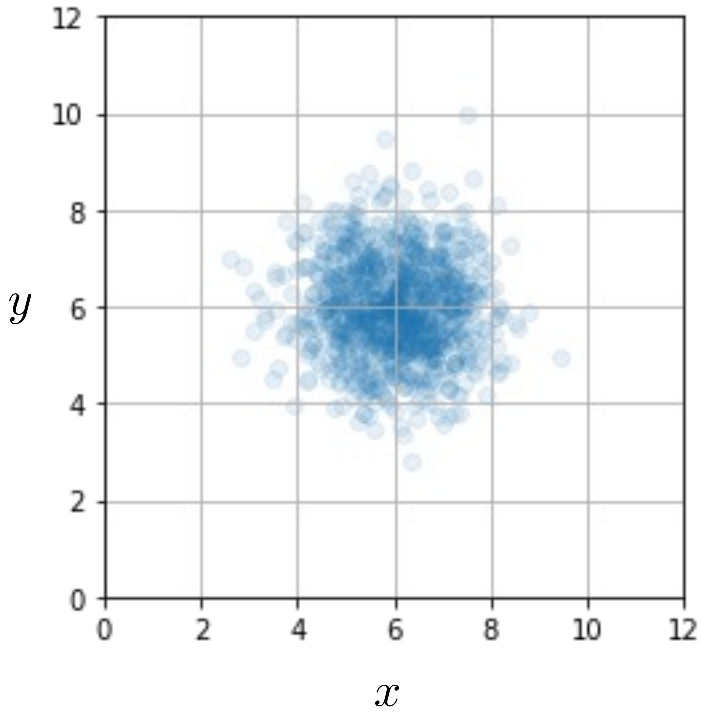
- *correlation coefficient* of two RVs:

$$\rho = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}}$$

Example

$(X, Y) \sim f(x, y)$



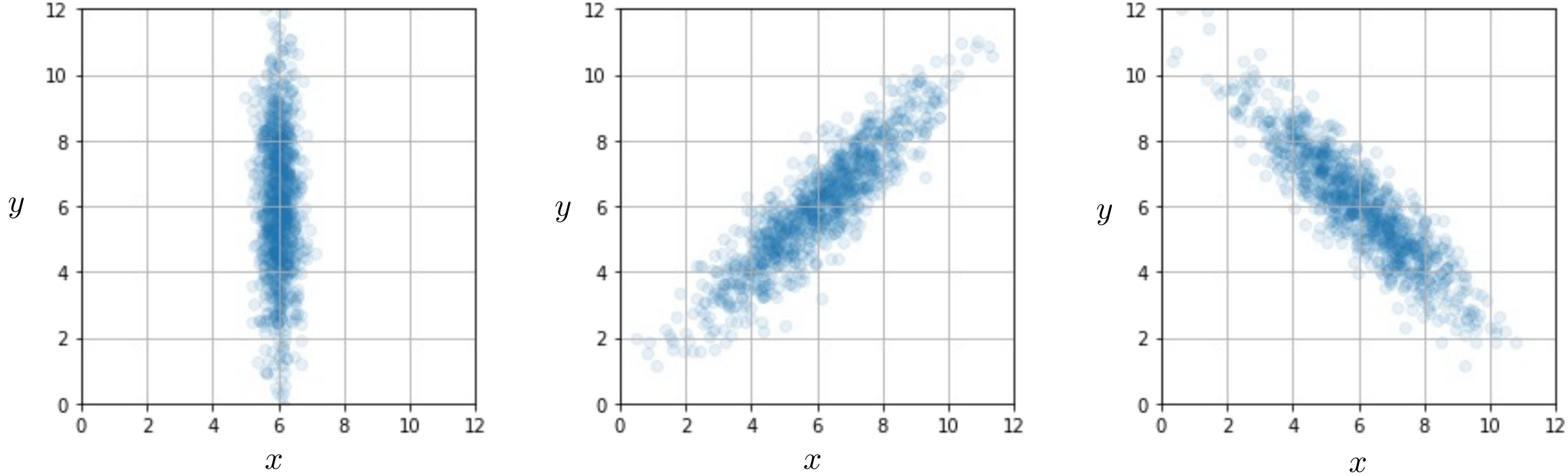


correlation?

$\mathrm{cov}(X, Y)$?

Example

$\mathrm{cov}(X, Y)$?



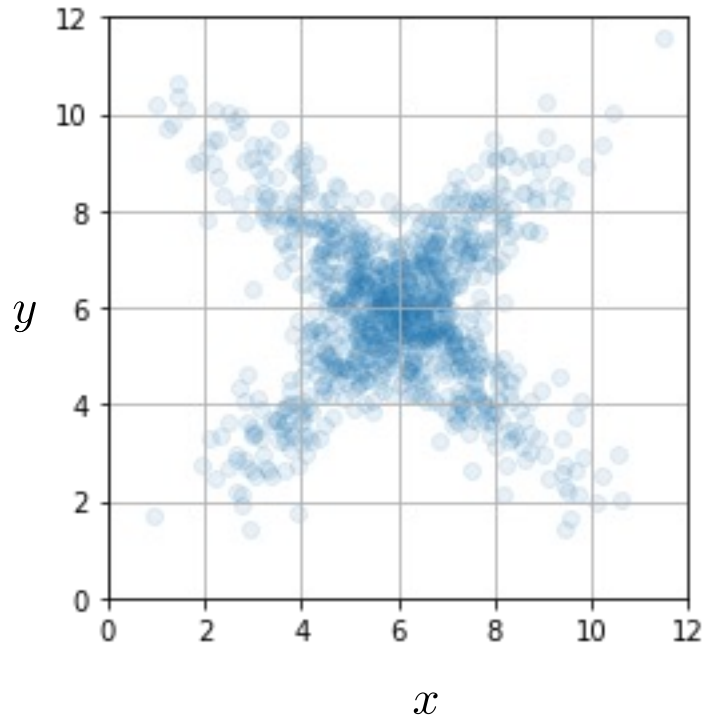
Properties

- if X, Y independent

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

$$\implies \mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y] = 0$$

- if $\mathrm{Cov}(X,Y) = 0$ then we say X, Y are *uncorrelated*
- *uncorrelated* does not imply independence



Intro to Random Vectors

Random Vectors

- consider random variables $X_1, X_2, \ldots, X_n$
- notation for random vectors: $\boldsymbol{x}, X, \boldsymbol{X}$ ...

$$\boldsymbol{x} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

- $\boldsymbol{x} \in \mathbb{R}^n$ is a random vector

Type equation here.

- a random vector $\boldsymbol{x}$ is specified by it joint cdf $F(\boldsymbol{x})$

$$F(\boldsymbol{x}) = \mathbb{P}(X_1 \leq x_1, X_2 \leq x_2, \ldots)$$

- if $\boldsymbol{x}$ is discrete, then completely specified by pmf:

$$p(\boldsymbol{x}) = p(x_1, x_2, \ldots) = \mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots)$$

- if $\boldsymbol{x}$ is continuous, then completely specified by pdf:

$$f(\boldsymbol{x}) = f(x_1, x_2 \ldots)$$

$$f(\boldsymbol{x}) = \frac{\partial^n F(\boldsymbol{x})}{\partial x_1 \cdots \partial x_n}$$

Marginals

- the marginal pdf, pmf or cdf is the joint distribution for a subset of the random variables

$$\boldsymbol{x} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}$$

- marginal of X_2, X_4?

$$f(x_2, x_4) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\boldsymbol{x}) dx_1 dx_3$$

$$p(x_2, x_4) = \sum_{x_1} \sum_{x_3} \quad p(\boldsymbol{x})$$

Mean Vector and Covariance Matrix

$$\boldsymbol{x} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

- mean of a random vector

$$\mu_x = E[\boldsymbol{x}] = \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix}$$

- covariance matrix of a random vector $\boldsymbol{x}$ is a matrix with covariance between all pairs of RVs

$$\Sigma_{\boldsymbol{x}} = E[(\boldsymbol{x} - \mu_x)(\boldsymbol{x} - \mu_x)^T]$$

$$\Sigma_{\boldsymbol{x}} = E[\boldsymbol{x}\boldsymbol{x}^T] - \mu_x \mu_x^T$$

$$\Sigma_{\boldsymbol{x}} = \begin{bmatrix} \mathrm{var}(X_1) & \mathrm{cov}(X_2, X_1) & \dots & \mathrm{cov}(X_1, X_n) \\ \mathrm{cov}(X_2, X_1) & \mathrm{var}(X_2) & \dots & \mathrm{cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(X_n, X_1) & \mathrm{cov}(X_n, X_2) & \dots & \mathrm{var}(X_n) \end{bmatrix}$$

Independence

$$\boldsymbol{x} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

- properties of $\Sigma_{\boldsymbol{x}}$:
  - Σ_X is real and symmetric
  - Σ_X is positive semi-definite,

    i.e, $\boldsymbol{a}^T \Sigma \boldsymbol{a} \geq 0$ for any $\boldsymbol{a}$.

- $X_1,\ X_2, \ldots, X_n$ are independent if and only if

  $p(x_1, x_2, \ldots, x_n) = p(x_1)p(x_2)\ldots p(x_n)$ for all $x_1, x_2 \ldots x_n \in \mathcal{X}^n$

  $p(\boldsymbol{x}) = \prod_{i=1}^{n} p(x_i)$

- if $X_1,\ X_2, \ldots, X_n$ are independent then

$$\Sigma_X = \begin{bmatrix} \mathrm{var}(X_1) & 0 & \ldots & 0 \\ 0 & \mathrm{var}(X_2) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \mathrm{var}(X_n) \end{bmatrix}$$