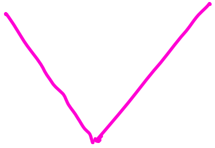
Spread of a random variable X
Mean: μ

$$\underbrace{\mathbb{E}\left[|X-\mu|\right]}_{\text{what we'd like}} \leq \underbrace{\sqrt{\mathbb{E}\left[(X-\mu)^2\right]}}_{\text{what we use (standard dev.)}}$$

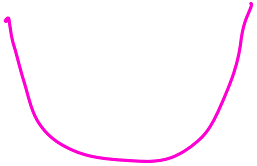
Modeling dependence between variables

DSE 210

$y = |x|$



$y = x^2$



Some r.v. X. Want to find a "central value" of X.

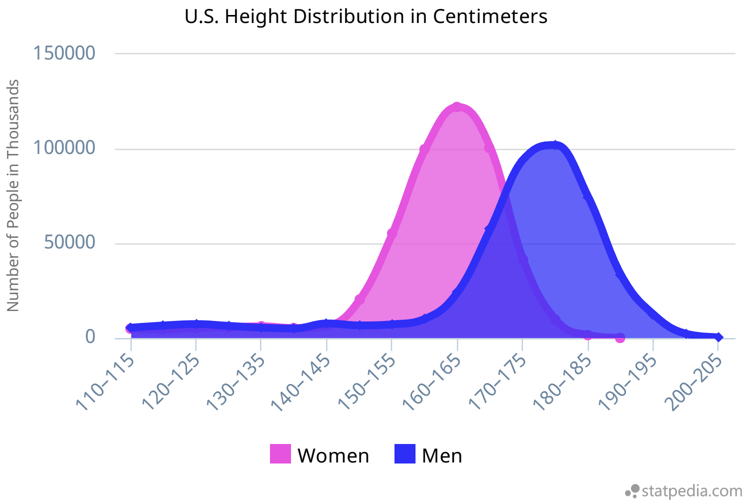
What value z minimizes

$\mathbb{E}\left[|X-z|\right]$? $\longrightarrow$ $z = \text{median}(X)$

$\mathbb{E}\left[(X-z)^2\right]$? $\longrightarrow$ $z = \text{mean}(X)$

Multiple random variables

We've seen many ways to model how a single variable, e.g. height, is distributed in a population.



What if we have more variables, e.g. weight as well?

Dependence

Example: For a person chosen at random from a population, take

two random variables $\left\{ \begin{array}{l} H = \text{height} \\ W = \text{weight} \end{array} \right.$

We could treat them as independent, e.g.

- Fit a Gaussian G_1 to the heights
- Fit a Gaussian G_2 to the weights

Independence would mean

$$\Pr(H = h, W = w) \;=\; \Pr(H = h)\Pr(W = w).$$

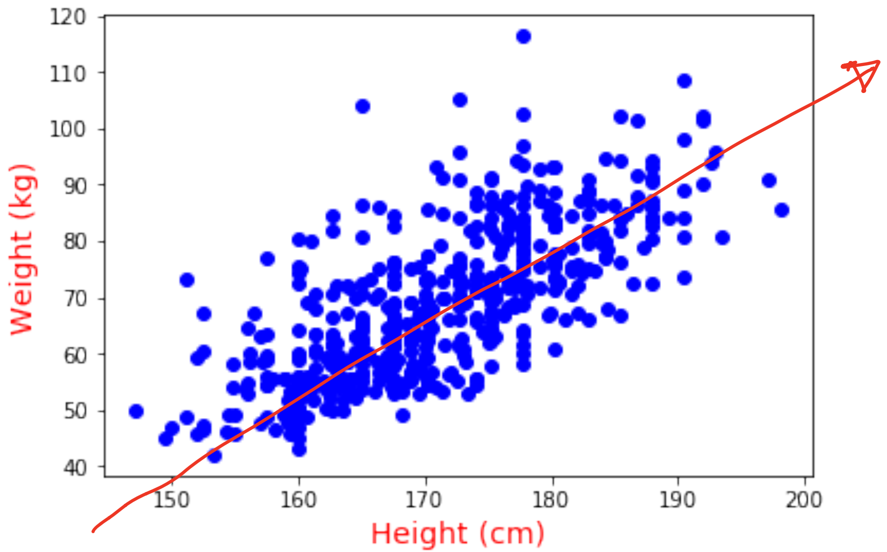
This is not a good model. Why? Height and weight are not independent.

Making an independence assumption leads to an inaccurate model of (H,W)

Correlation

When H gets higher,
W also gets higher <u>on average</u>.

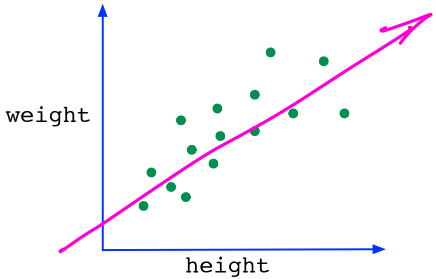
Height and weight are **positively correlated**.



general upward trend

Based on body measurements of 507 people at `https://ww2.amstat.org/publications/jse/datasets/body.txt`

Types of correlation



$H, W \in \{1,2,3\}$

$\mathbb{E}H = 2$

$\mathbb{E}W = 2$

H	W	
1	1	1
2	2	4
3	3	9

$\rightarrow \mathbb{E}[HW] =$

H, W **positively correlated**

This also implies

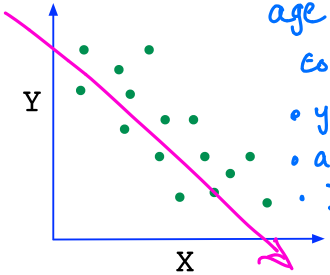
$$\underbrace{\mathbb{E}[HW]}_{4\,2/3} > \underbrace{\mathbb{E}[H]\,\mathbb{E}[W]}_{4}$$

e.g. age is positively correlated with:

- risk of cancer, heart failure
- hearing loss
- wisdom

H	W	HW
1	3	3
2	2	4
3	1	3

$\mathbb{E}[HW] = 3\,1/3$

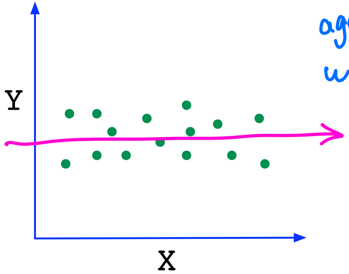


age is negatively correlated with:

- youthfulness
- amt of hair
- ID checks

X, Y **negatively correlated**

$$\mathbb{E}[XY] < \mathbb{E}[X]\,\mathbb{E}[Y]$$



age is uncorrelated with:

- weather
- phone number
- SS #

X, Y **uncorrelated**

$$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y]$$

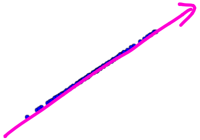
Correlation coefficient: pictures

$r \in [-1, 1]$

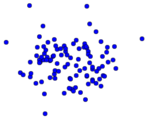
perfect negative correlation

perfect positive corr

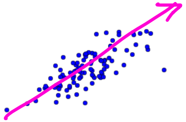
$r = 1$



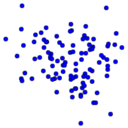
$r = 0$



$r = 0.75$



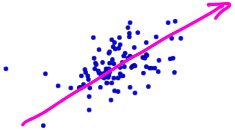
$r = -0.25$



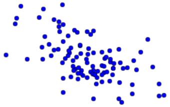
$|r| > 0.5$

fairly strong correlation

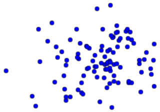
$r = 0.5$



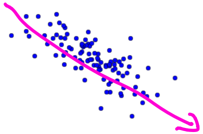
$r = -0.5$



$r = 0.25$



$r = -0.75$



Covariance and correlation

pos. correlation $\Leftrightarrow$ $\mathbb{E}[XY] > \mathbb{E}[X]\mathbb{E}[Y]$ $\Leftrightarrow$ cov > 0

neg. correlation $\Leftrightarrow$ $\mathbb{E}[XY] < \mathbb{E}[X]\mathbb{E}[Y]$ $\Leftrightarrow$ cov < 0

uncorrelated $\Leftrightarrow$ $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ $\Leftrightarrow$ cov = 0

- **Covariance**

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

$$\mathrm{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$$

Maximized when $X = Y$, in which case it is $\mathrm{var}(X)$.
In general, it is at most $\mathrm{std}(X)\mathrm{std}(Y)$.

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

linearity of expectation

- **Correlation**

$$\mathrm{corr}(X, Y) = \frac{\mathrm{cov}(X, Y)}{\mathrm{std}(X)\mathrm{std}(Y)}$$

This is always in the range $[-1, 1]$.

X, Y indpt $\Rightarrow$ $\mathrm{corr}(X,Y) = 0$

$\mathrm{corr}(X,Y) = 0$ $\not\Rightarrow$ X, Y indpt

Example 1

Find cov(X, Y) and corr(X, Y)

"Joint distribution"

x	y	$\Pr(x,y)$
-1	-1	$1/3$
-1	1	$1/6$
1	-1	$1/3$
1	1	$1/6$

$$\mathrm{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

$$\mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\mathrm{std}(X)\,\mathrm{std}(Y)}$$

xy	Pr
-1	$1/2$
1	$1/2$

$\mathbb{E}[XY] = 0$

x	Pr
-1	$1/2$
1	$1/2$

y	Pr
-1	$2/3$
1	$1/3$

$\mathbb{E}[X] = 0$

$\mathbb{E}[X^2] = 1$

$\mathrm{var}(X) = 1$

$\mathrm{std}(X) = 1$

$\mathbb{E}[Y] = -1/3$

$\mathbb{E}[Y^2] = 1$

$\mathrm{var}(Y) = 8/9$

$\mathrm{std}(Y) = \sqrt{8/9}$

$\mathrm{cov}(X,Y) = 0$

$\mathrm{corr}(X,Y) = 0$

Are X, Y ind$_p$t?

Yes – by checking all entries in the joint distribution table for X, Y

Example 2

Find cov(X, Y) and corr(X, Y)

x	y	$\Pr(x,y)$
-1	-10	$1/6$
-1	10	$1/3$
1	-10	$1/3$
1	10	$1/6$

x	Pr
-1	1/2
1	1/2

y	Pr
-10	1/2
10	1/2

$\mathbb{E}[X] = 0$

$\text{var}(X) = 1$

$\text{std}(X) = 1$

$\mathbb{E}[Y] = 0$

$\text{var}(Y) = 100$

$\text{std}(Y) = 10$

xy	Pr
-10	2/3
10	1/3

$\Big\}\ \mathbb{E}[XY] = -10/3$

$$\text{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] = -10/3$$

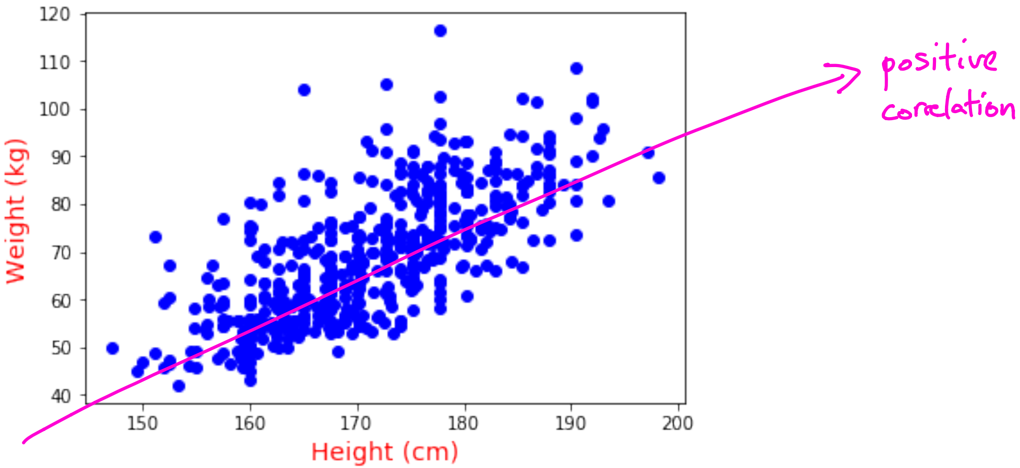
$$\text{corr}(X,Y) = \frac{\text{cov}(X,Y)}{\text{std}(X)\,\text{std}(Y)} = \frac{-10/3}{1 \cdot 10} = -\frac{1}{3}$$

Are X, Y indpt?

No – because they are negatively correlated

(no need to check table!)

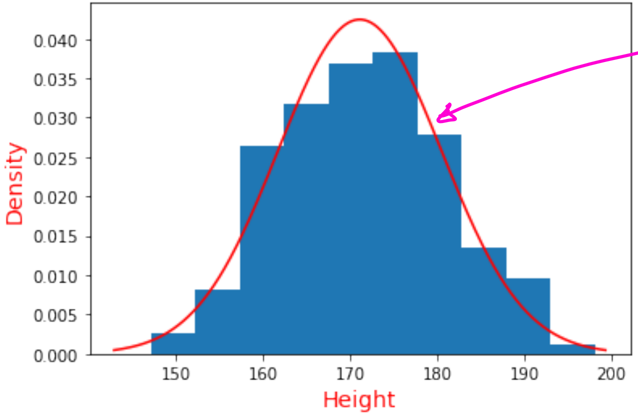
Height and weight again



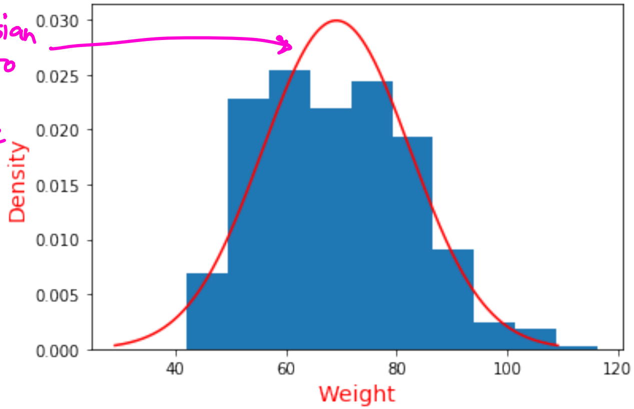
- Height (cm): $\mathbb{E}(H) = 171.1$, $\text{std}(H) = 9.4$
- Weight (kg): $\mathbb{E}(W) = 69.1$, $\text{std}(W) = 13.3$
- $\mathbb{E}(HW) = 11924.0$ while $\mathbb{E}(H)\mathbb{E}(W) = 11834.2$
- $\text{cov}(H, W) = 89.9$ and $\text{corr}(H, W) = 0.72$ strong correlation

A distribution over two variables?

We want a distribution over two variables: $(X_1, X_2) = (\text{height}, \text{weight})$



Gaussian fit to each variable



- Mean $\mu_1 = 171.1$
- Standard dev $\sigma_1 = 9.4$

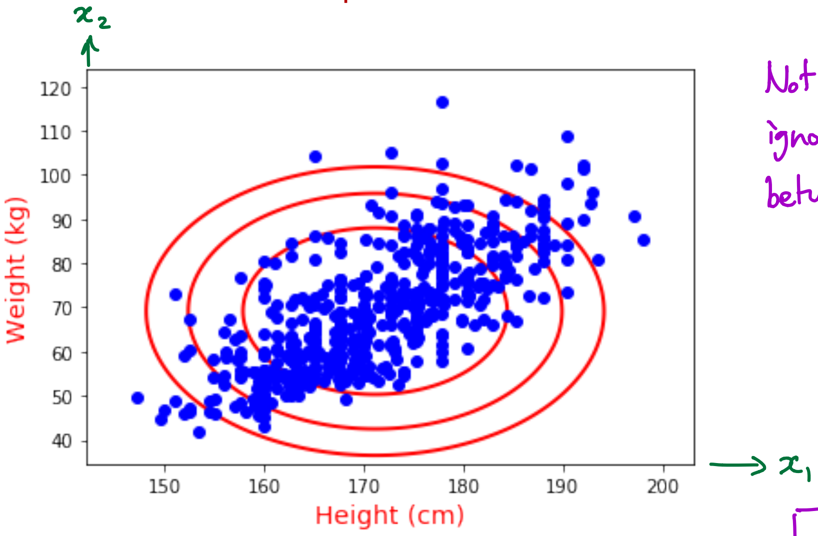
- Mean $\mu_1 = 69.1$
- Standard dev $\sigma_1 = 13.3$

Independent variables

One possibility: Treat the two variables as independent and fit a Gaussian to each.

$X_1 \sim N(\underline{\mu_1}, \underline{\sigma_1^2})$

$X_2 \sim N(\underline{\mu_2}, \underline{\sigma_2^2})$



Not a good fit: ignores correlation between X_1 and X_2.

density over $\mathbb{R}^2$

What is the resulting density over (x_1, x_2)?

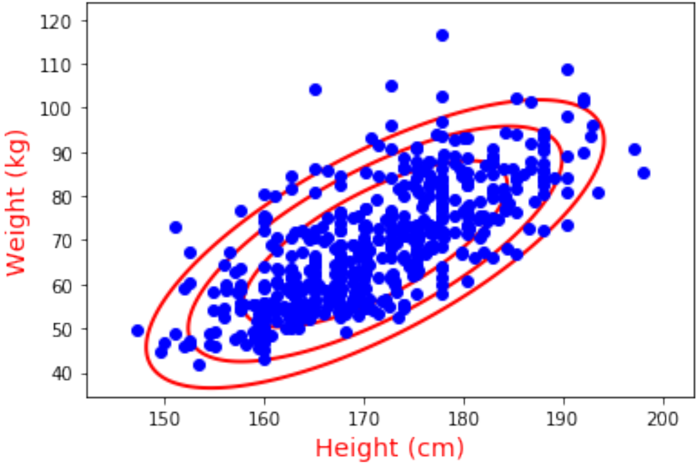
$$p(x_1,x_2) = p_1(x_1)\, p_2(x_2) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-(x_1-\mu_1)^2/2\sigma_1^2} \cdot \frac{1}{\sigma_2\sqrt{2\pi}} e^{-(x_2-\mu_2)^2/2\sigma_2^2}$$

Four params: $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$

The bivariate Gaussian

(2-dimensional)

$\boxed{\text{Include a } \underline{\text{fifth}} \text{ parameter : } \mathrm{cov}(X_1, X_2)}$

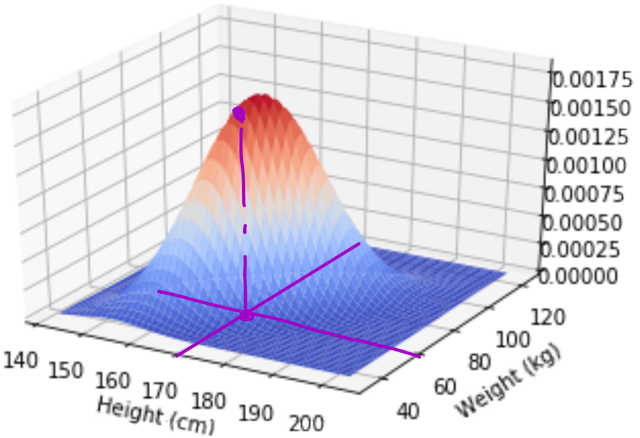


Much better fit.

Model the data by a bivariate Gaussian, parametrized by:

$$\text{mean } \mu = \begin{pmatrix} 171.1 \\ 69.1 \end{pmatrix} \begin{matrix} \mu_1 \\ \mu_2 \end{matrix} \quad \text{and covariance matrix } \Sigma = \begin{pmatrix} \overset{\sigma_1^2}{88.4} & \overset{\mathrm{cov}(x_1,x_2)}{89.9} \\ 89.9 & \underset{\sigma_2^2}{176.9} \end{pmatrix}$$

The bivariate Gaussian



Model the data by a bivariate Gaussian, parametrized by:

$$\text{mean } \mu = \begin{pmatrix} 171.1 \\ 69.1 \end{pmatrix} \text{ and covariance matrix } \Sigma = \begin{pmatrix} 88.4 & 89.9 \\ 89.9 & 176.9 \end{pmatrix}$$

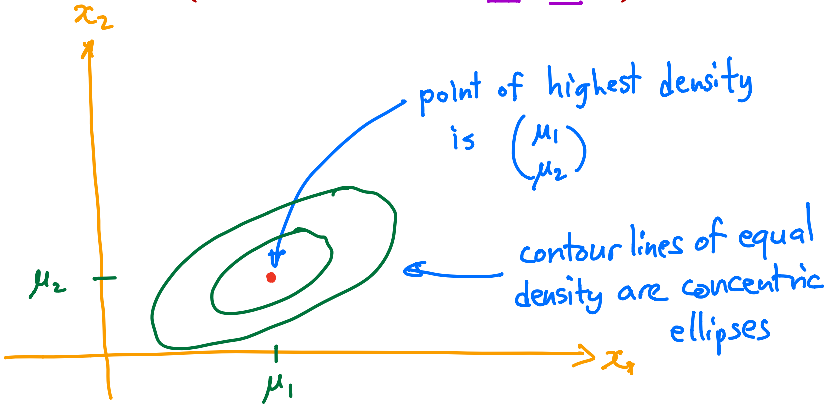
The bivariate (2-d) Gaussian

A distribution over $(x_1, x_2) \in \mathbb{R}^2$, parametrized by:

- **Mean** $(\mu_1, \mu_2) \in \mathbb{R}^2$, where $\mu_1 = \mathbb{E}(X_1)$ and $\mu_2 = \mathbb{E}(X_2)$
- **Covariance matrix** $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ where $\left\{ \begin{array}{c} \Sigma_{11} = \text{var}(X_1) \\ \Sigma_{22} = \text{var}(X_2) \\ \Sigma_{12} = \Sigma_{21} = \text{cov}(X_1, X_2) \end{array} \right\}$

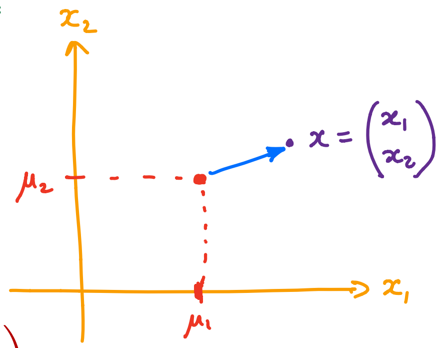
$$\Sigma = \begin{bmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) \\ \text{cov}(X_1, X_2) & \text{var}(X_2) \end{bmatrix}$$

Density is highest at the mean, falls off in ellipsoidal contours.



Density of the bivariate Gaussian

$\exp(-z) = e^{-z}$



- **Mean** $(\mu_1, \mu_2) \in \mathbb{R}^2$, where $\mu_1 = \mathbb{E}(X_1)$ and $\mu_2 = \mathbb{E}(X_2)$
- **Covariance matrix** $\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$

Density $$p(x_1, x_2) = \frac{1}{2\pi \underbrace{|\Sigma|}_{\text{determinant of matrix } \Sigma}^{1/2}} \exp\left(-\frac{1}{2} \underbrace{\begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}^T}_{\text{2d-vector}} \Sigma^{-1} \underbrace{\begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}}_{\text{2d-vector}} \right)$$

inverse of Σ is also a 2x2 matrix

$$\underbrace{\begin{pmatrix} \text{height} - \text{avg. height} \\ \text{weight} - \text{avg. weight} \end{pmatrix}}_{\begin{pmatrix} 2 \\ 1 \end{pmatrix}}$$

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$$

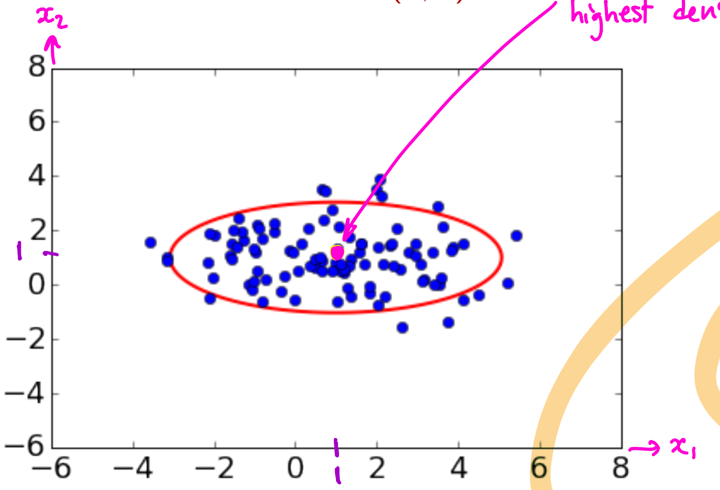
E.g. $\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$

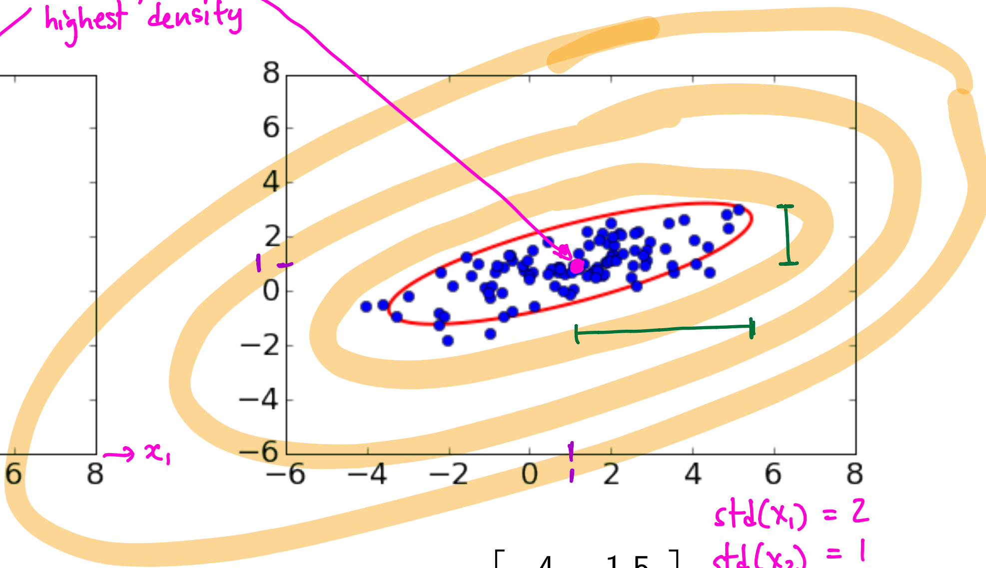
$$(2 \quad 1) \; \Sigma^{-1} \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

Bivariate Gaussian: examples

In either case, the mean is $(1, 1)$.

mean: point of highest density





std $(x_1) = 2$

std $(x_2) = 1$

$\text{cov}(x_1, x_2) = 0$

$\text{corr}(x_1, x_2) = 0$

$$\Sigma = \begin{array}{ccc} & \text{var}(x_1) & \\ & \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} & \\ & & \text{var}(x_2) \\ & \text{cov}(x_1, x_2) & \end{array}$$

$$\Sigma = \begin{bmatrix} 4 & 1.5 \\ 1.5 & 1 \end{bmatrix}$$

std$(x_1) = 2$

std$(x_2) = 1$

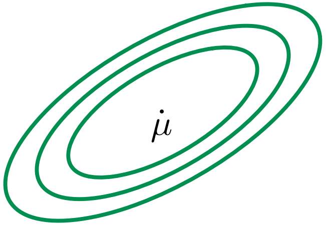
$\text{cov}(x_1, x_2) = 1.5$

$\text{corr}(x_1, x_2) = 0.75$

The multivariate Gaussian

Move to data with d variables.

E.g. $d=3$: $\begin{array}{ccc} (X_1, & X_2, & X_3) \\ \text{height} & \text{weight} & \text{age} \end{array}$



$N(\mu, \Sigma)$: Gaussian in $\mathbb{R}^d$

- mean: $\mu \in \mathbb{R}^d$ $\leftarrow$ $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}$ $\begin{array}{l} \leftarrow \text{mean height} \\ \leftarrow \text{mean weight} \\ \leftarrow \text{mean age} \end{array}$
- covariance: $d \times d$ matrix Σ

Generates points $X = (X_1, X_2, \ldots, X_d)$.

covariance matrix Σ $\longrightarrow$ $\begin{bmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) & \text{cov}(X_1, X_3) \\ \text{cov}(X_2, X_1) & \text{var}(X_2) & \text{cov}(X_2, X_3) \\ \text{cov}(X_3, X_1) & \text{cov}(X_3, X_2) & \text{var}(X_3) \end{bmatrix}$

- μ is the vector of coordinate-wise means:

$$\mu_1 = \mathbb{E}X_1, \ \mu_2 = \mathbb{E}X_2, \ldots, \ \mu_d = \mathbb{E}X_d.$$

- Σ is a matrix containing all pairwise covariances:

$$\Sigma_{ij} = \Sigma_{ji} = \text{cov}(X_i, X_j) \quad \text{if } i \neq j$$
$$\Sigma_{ii} = \text{var}(X_i)$$

inverse of cov. matrix

Density $\quad p(x) = \dfrac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\dfrac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$

determinant

Special case: independent features

$$\underbrace{(X_1, X_2, \ldots, X_d)}_{\text{actually indpt}}$$

Suppose the X_i are independent, and $\text{var}(X_i) = \sigma_i^2$.

What is the covariance matrix Σ, and what is its inverse Σ^{-1}?

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

diagonal matrix

For diagonal matrices, it's easy to write down the inverse and determinant.

$$\Sigma^{-1} = \begin{bmatrix} 1/\sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & 1/\sigma_d^2 \end{bmatrix}$$

$$|\Sigma| = \sigma_1^2 \sigma_2^2 \cdots \sigma_d^2$$

Diagonal Gaussian

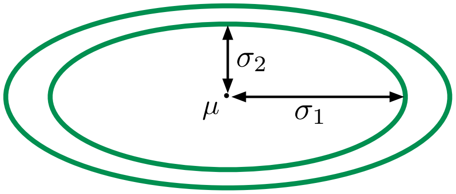
Diagonal Gaussian: the X_i are independent, with variances σ_i^2. Thus

$$\Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_d^2) \quad \text{(off-diagonal elements zero)}$$

Each X_i is an independent one-dimensional Gaussian $N(\mu_i, \sigma_i^2)$:

$$\Pr(x) = \Pr(x_1)\Pr(x_2)\cdots\Pr(x_d) = \frac{1}{(2\pi)^{d/2}\sigma_1\cdots\sigma_d}\exp\left(-\sum_{i=1}^{d}\frac{(x_i-\mu_i)^2}{2\sigma_i^2}\right)$$

Contours of equal density are **axis-aligned ellipsoids** centered at μ:



How to fit a Gaussian to data

Fit a Gaussian to data points $x^{(1)}, \ldots, x^{(m)} \in \mathbb{R}^d$.

- Empirical mean

$$\mu = \frac{1}{m}\left(x^{(1)} + \cdots + x^{(m)}\right)$$

mean of the data vectors

- Empirical covariance matrix has i, j entry:

$$\Sigma_{ij} = \left(\frac{1}{m}\sum_{k=1}^{m} x_i^{(k)} x_j^{(k)}\right) - \mu_i \mu_j$$

all covariances