Gaussian Random Vectors

Probability and Statistics for Data Science

Carlos Fernandez-Granda

NYU | COURANT INSTITUTE OF MATHEMATICAL SCIENCES

NYU DATA SCIENCE

These slides are based on the book Probability and Statistics for Data Science by Carlos Fernandez-Granda, available for purchase here. A free preprint, videos, code, slides and solutions to exercises are available at `https://www.ps4ds.net`

Goal

Define Gaussian parametric model for random vectors

Motivation

Curse of dimensionality

Nonparametric density estimation is impossible in high dimensions

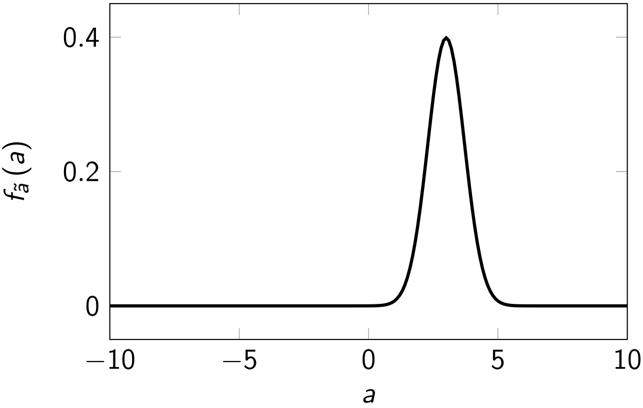
Gaussian distribution

Motivation: Sum of independent quantities is approximately Gaussian

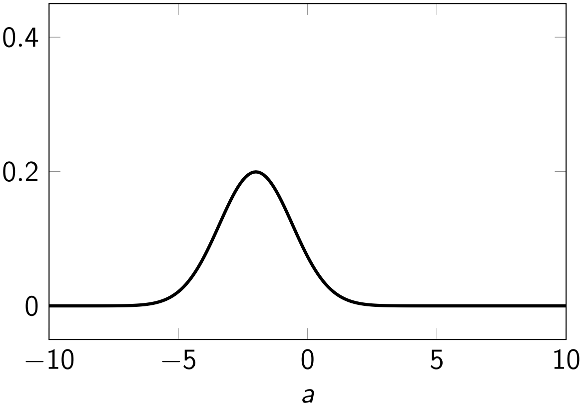
The Gaussian or normal parametric pdf with mean μ and standard deviation σ is

$$f_{\tilde{a}}(a) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(a-\mu)^2}{2\sigma^2}}$$

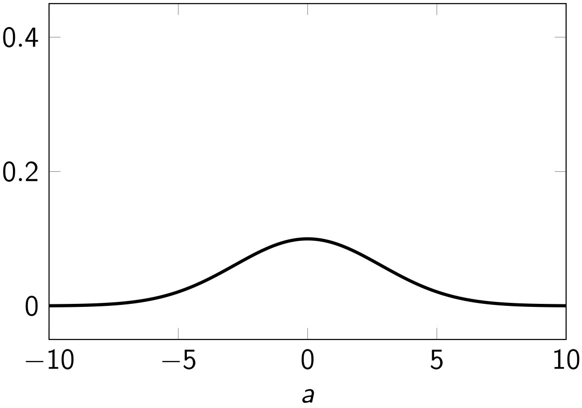
$\mu = 3$, $\sigma = 1$



$\mu = -2$, $\sigma = 2$



$\mu = 0,\ \sigma = 4$



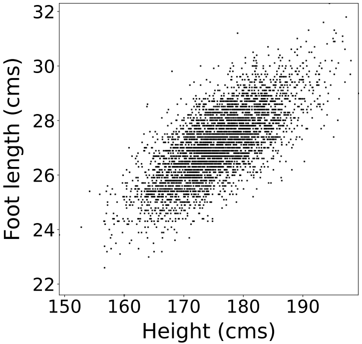
First try

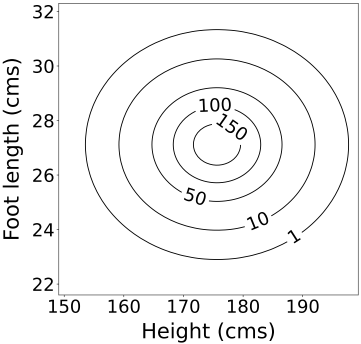
Entries of $\tilde{x}$ are d independent Gaussian random variables with means μ_1, μ_2, ..., μ_d and standard deviations σ_1, σ_2, ..., σ_d

Joint pdf

$$
\begin{aligned}
f_{\tilde{x}}(x) &= \prod_{i=1}^{d} f_{\tilde{x}[i]}\left(\tilde{x}[i]\right) \\
&= \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{(x[i]-\mu_i)^2}{2\sigma_i^2}\right) \\
&= \frac{1}{(2\pi)^{\frac{d}{2}} \prod_{i=1}^{d} \sigma_i} \exp\left(-\frac{1}{2}\sum_{i=1}^{d} \frac{(x[i]-\mu_i)^2}{\sigma_i^2}\right)
\end{aligned}
$$

Height and foot length





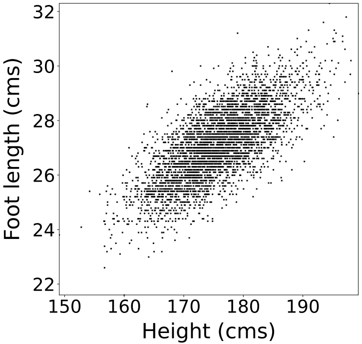
Contour surfaces

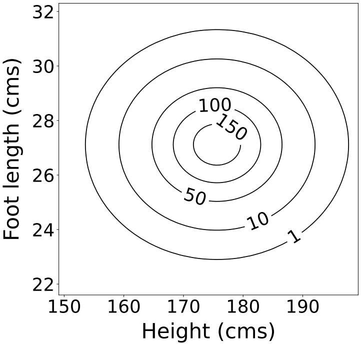
$$\left\{x \in \mathbb{R}^d \mid f_{\tilde{x}}(x) = c\right\} = \left\{x \in \mathbb{R}^d \mid \sum_{i=1}^{d} \frac{(x[i] - \mu_i)^2}{\sigma_i^2} = c'\right\}$$

where $c' = -2 \log \left(c \, (2\pi)^{\frac{d}{2}} \prod_{i=1}^{d} \sigma_i \right)$

Ellipsoid with axes along coordinate axes

The model is too rigid!





Including rotations

Additional parameters: Axes of ellipsoid u_1, u_2, ..., u_d

$$\begin{aligned} c' &= \sum_{i=1}^{d} \frac{{u_i}^T (x-\mu)^2}{\sigma_i^2} \\ &= (x-\mu)^T U \Lambda^{-1} U^T (x-\mu) \\ &= (x-\mu)^T \Sigma^{-1} (x-\mu) \end{aligned}$$

$$U := \begin{bmatrix} u_1 & u_2 & \cdots & u_d \end{bmatrix} \qquad \Lambda := \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \sigma_d^2 \end{bmatrix}$$

Covariance-matrix parameter $\Sigma := U \Lambda U^T$

Joint pdf

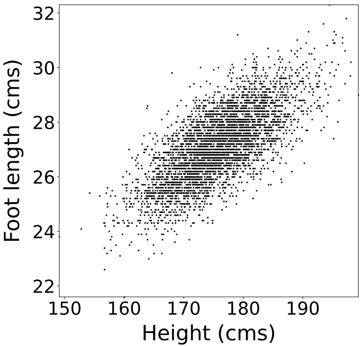
Without rotation:

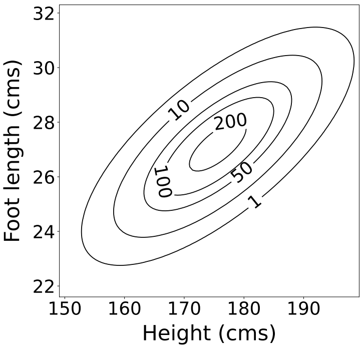
$$f_{\tilde{x}}(x) = \frac{1}{(2\pi)^{\frac{d}{2}} \prod_{i=1}^{d} \sigma_i} \exp\left(-\frac{1}{2} \sum_{i=1}^{d} \frac{(x[i] - \mu_i)^2}{\sigma_i^2} \right)$$

With rotation:

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

With rotation





Gaussian random vector

A Gaussian random vector $\tilde{x}$ is a random vector with joint pdf

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

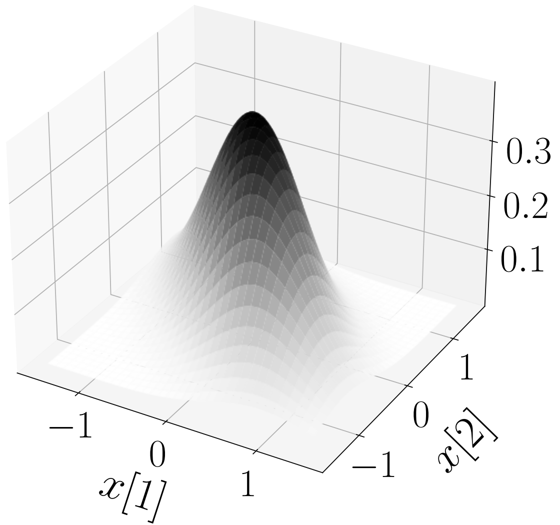
where $\mu \in \mathbb{R}^d$ is the mean and $\Sigma \in \mathbb{R}^{d\times d}$ the covariance matrix

$\Sigma \in \mathbb{R}^{d\times d}$ is symmetric and positive definite (positive eigenvalues)

2D example

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 0.5 & -0.3 \\ -0.3 & 0.5 \end{bmatrix}$$

Density



2D example

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 0.5 & -0.3 \\ -0.3 & 0.5 \end{bmatrix}$$

How do the contour lines look like?

Spectral theorem

If $A \in \mathbb{R}^{d \times d}$ is symmetric, then it has an eigendecomposition

$$\begin{aligned} A &= U \Lambda U^T \\ &= \begin{bmatrix} u_1 & u_2 & \cdots & u_d \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ & & \cdots & \\ 0 & 0 & \cdots & \lambda_d \end{bmatrix} \begin{bmatrix} u_1 & u_2 & \cdots & u_d \end{bmatrix}^T \end{aligned}$$

Eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$ are real

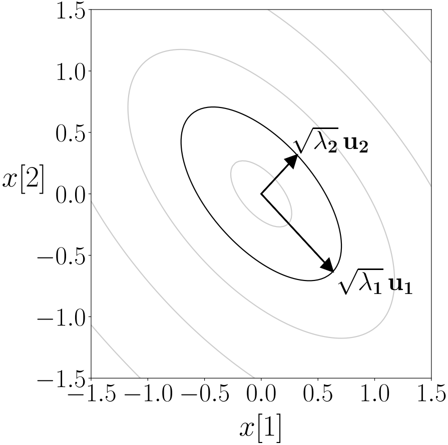
Eigenvectors $u_1, u_2, \ldots, u_n$ are real and orthogonal

Contour surfaces

$$
\begin{aligned}
c &= x^T \Sigma^{-1} x \\
&= x^T U \Lambda^{-1} U^T x \\
&= \sum_{i=1}^{d} \frac{(u_i^T x)^2}{\lambda_i}
\end{aligned}
$$

Ellipsoid with axes proportional to $\sqrt{\lambda_i}$

Contour surfaces



Log-likelihood

Data: $X := \{x_1, \dots, x_n\}$

Log-likelihood of Gaussian parameters

$$\log \mathcal{L}(\mu, \Sigma) := \log \prod_{i=1}^{n} \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left(-\frac{1}{2}(x_i - \mu)^T \Sigma^{-1} (x_i - \mu)\right)$$

Log-likelihood

$$
\begin{aligned}
&\arg\max_{\mu,\Sigma} \log \mathcal{L}(\mu, \Sigma) \\
&= \arg\max_{\mu,\Sigma} \log \prod_{i=1}^{n} \frac{1}{\sqrt{(2\pi)^d \left|\Sigma\right|}} \exp\left(-\frac{1}{2} \left(x_i - \mu\right)^T \Sigma^{-1} \left(x_i - \mu\right)\right) \\
&= \arg\max_{\mu,\Sigma} -\frac{1}{2} \sum_{i=1}^{n} \left(x_i - \mu\right)^T \Sigma^{-1} \left(x_i - \mu\right) - \frac{n}{2} \log \left|\Sigma\right| \\
&= \arg\min_{\mu,\Sigma} \frac{1}{2} \sum_{i=1}^{n} \left(x_i - \mu\right)^T \Sigma^{-1} \left(x_i - \mu\right) + \frac{n}{2} \log \left|\Sigma\right| \\
&= \arg\min_{\mu,\Sigma} g(\mu, \Sigma)
\end{aligned}
$$

Mean parameter

$$\nabla_\mu g(\mu, \Sigma) = \Sigma^{-1} \sum_{i=1}^{n} (x_i - \mu)$$

Σ^{-1} is positive definite by assumption, so this quantity can only be zero if

$$\sum_{i=1}^{n} (x_i - \mu) = 0$$

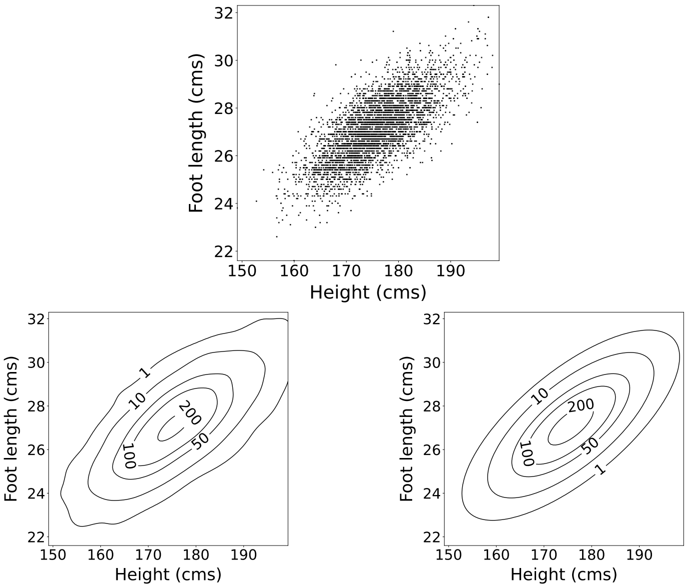
$$\mu_{\mathrm{ML}} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Maximum likelihood

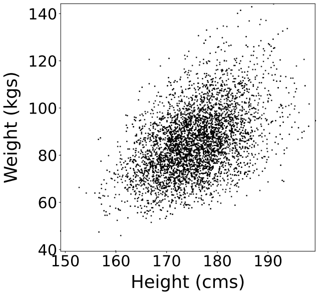
$$\mu_{\text{ML}} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

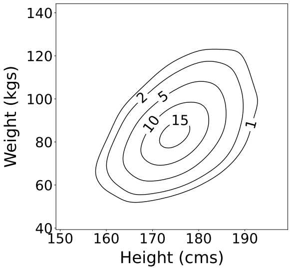
$$\Sigma_{\text{ML}} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_{\text{ML}})(x_i - \mu_{\text{ML}})^T$$

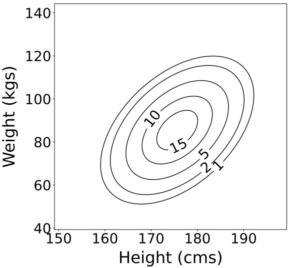
Height and foot length



Height and weight







What have we learned?

Definition of Gaussian random vectors

Analysis of contour surfaces

Maximum-likelihood parameter estimation