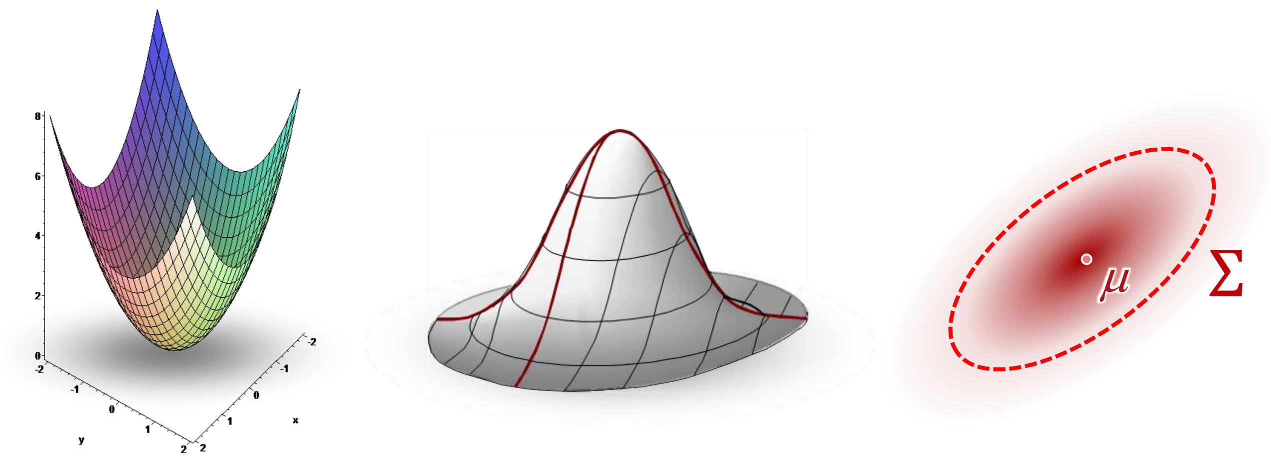
Modelling 1

SUMMER TERM 2020



LECTURE 12

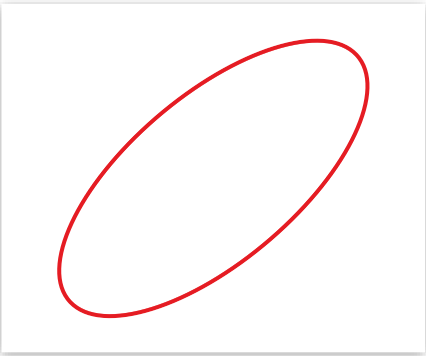
Gaussians

Michael Wand · Institut für Informatik · Michael.Wand@uni-mainz.de

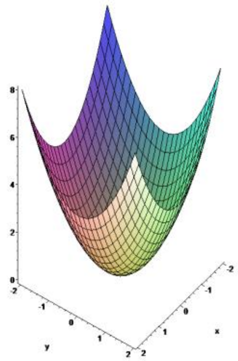
Quadratic Forms

The Iso-Lines: Quadrics

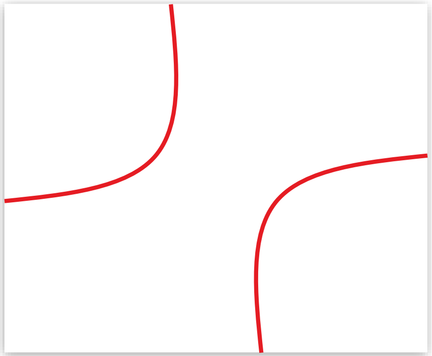
elliptic



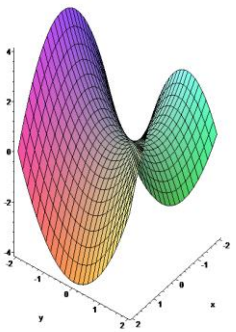
$\lambda_1 > 0, \lambda_2 > 0$



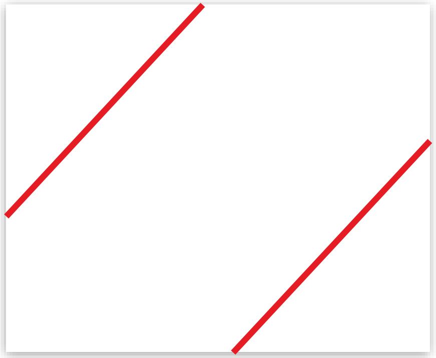
hyperbolic



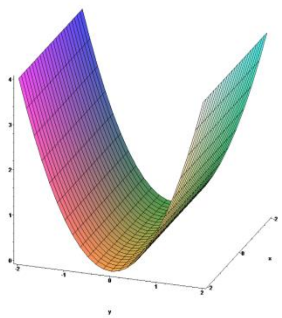
$\lambda_1 < 0, \lambda_2 > 0$



degenerate case

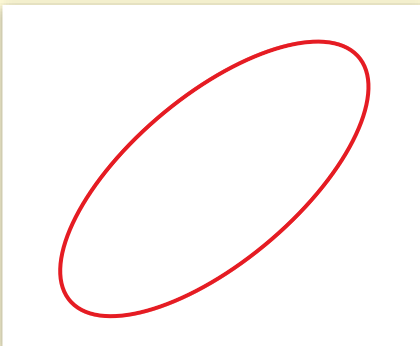


$\lambda_1 = 0, \lambda_2 \neq 0$

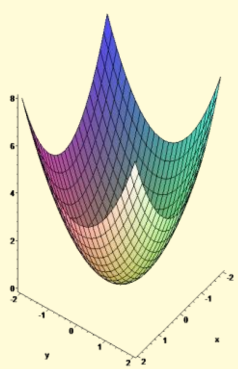


SPD Quadrics

elliptic

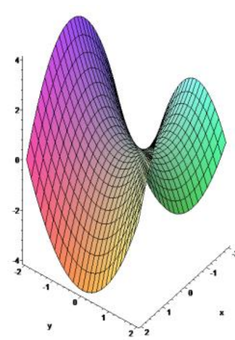


$\lambda_1 > 0, \lambda_2 > 0$

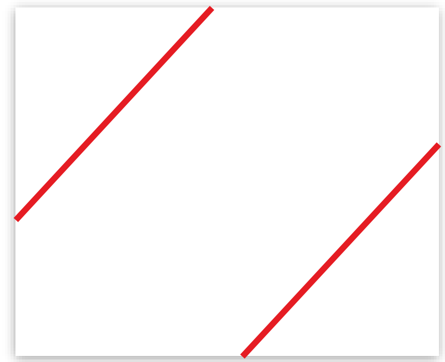


hyperbolic

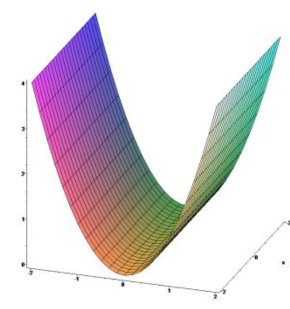
$\lambda_1 < 0, \lambda_2 > 0$



degenerate case



$\lambda_1 = 0, \lambda_2 \neq 0$



Gaussians

Gaussians

Gaussian Normal Distribution

- Two parameters: μ, σ
- Density:

$$\mathcal{N}_{\mu,\sigma}(x) := \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Mean: μ
- Variance: σ^2

Gaussian normal distribution

Log Space

Neg-log-density

$$\log \mathcal{N}_{\mu,\sigma}(x) := \frac{(x-\mu)^2}{2\sigma^2} + \frac{1}{2}\ln(2\pi\sigma^2)$$
$$= \frac{1}{2\sigma^2}(x-\mu)^2 + \text{const.}$$

Calculations in log-space

- Densities of products of Gaussians are Sums of quadratic polynomials
- Calculations simplified in log-space
  - Attention: Sum of Gaussians do not simplify!

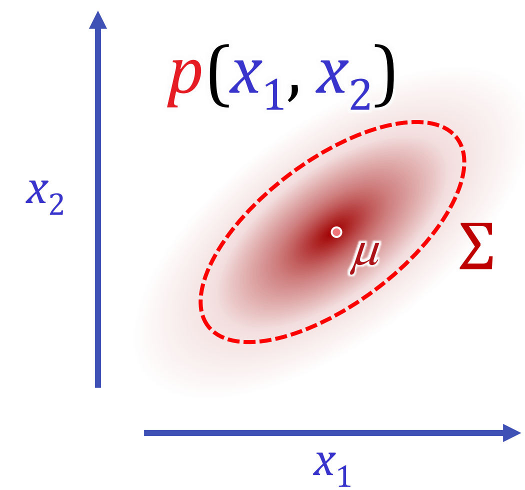
Multi-Variate Gaussians

Gaussian Normal Distribution in *d* Dimensions

- Two parameters: $\boldsymbol{\mu}$ (d-dim-vector), Σ ($d \times d$ matrix)
- Density:

$$\mathcal{N}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{x}) \coloneqq \left(\frac{1}{(2\pi)^{-\frac{d}{2}} \det(\Sigma)^{-\frac{1}{2}}}\right) e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

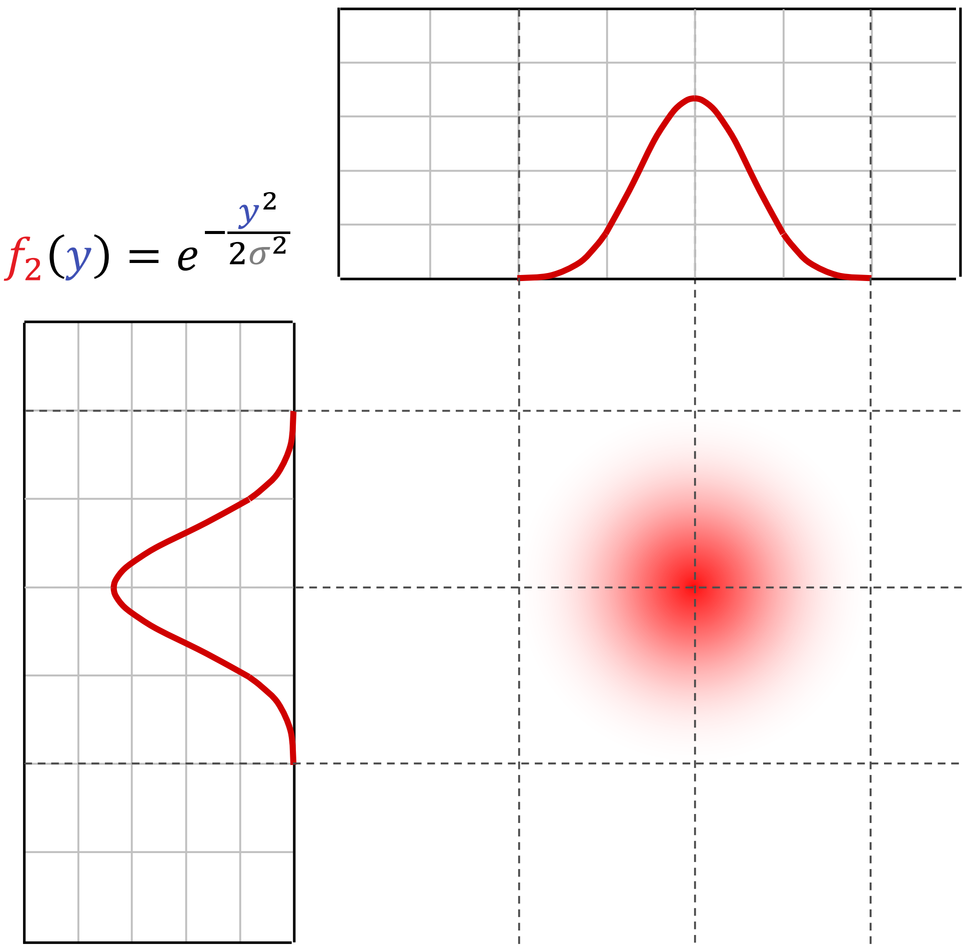
- Mean: $\boldsymbol{\mu}$
- Covariance Matrix: Σ



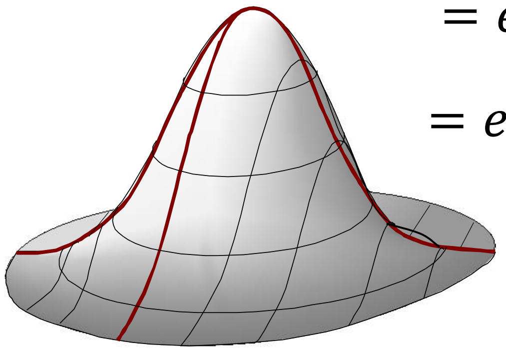
Factorization

$$f_1(x) = e^{-\frac{x^2}{2\sigma^2}}$$

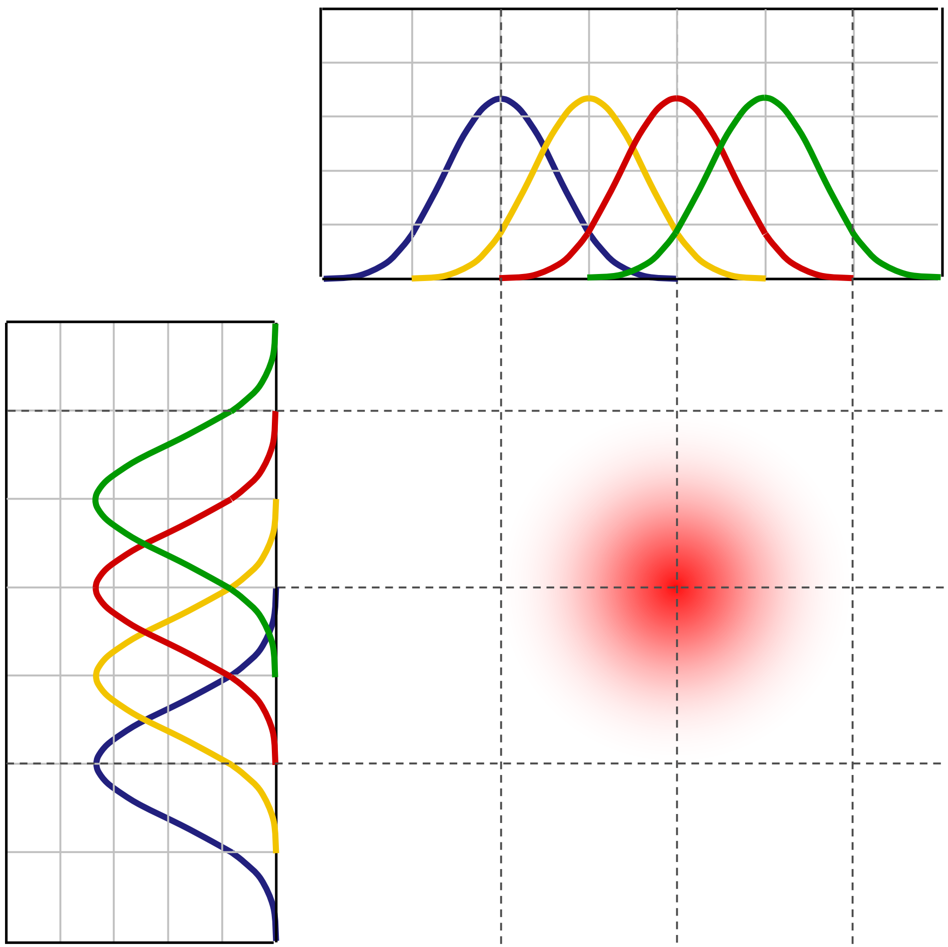
$$f_2(y) = e^{-\frac{y^2}{2\sigma^2}}$$



$$\begin{aligned} f(x,y) &= f_1(x) f_2(y) \\ &= e^{-\frac{x^2}{2\sigma^2}} \cdot e^{-\frac{y^2}{2\sigma^2}} \\ &= e^{-\frac{x^2+y^2}{2\sigma^2}} \\ &= e^{-\frac{1}{2}\mathbf{x}^{\mathrm{T}}\begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{pmatrix}\mathbf{x}} \end{aligned}$$

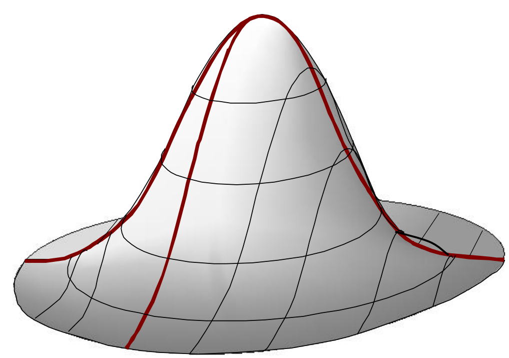


Remark: Tensor-Product Basis



$$\{f_1(x), f_2(x), \ldots, f_n(x)\}$$

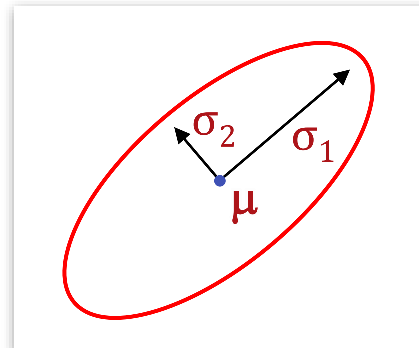
$$\rightarrow \left\{ \begin{matrix} f_1(x)f_1(y), & \cdots & f_n(x)f_1(y), \\ \vdots & & \vdots \\ f_1(x)f_n(y), & \cdots & f_n(x)f_n(y) \end{matrix} \right\}$$



Log Space

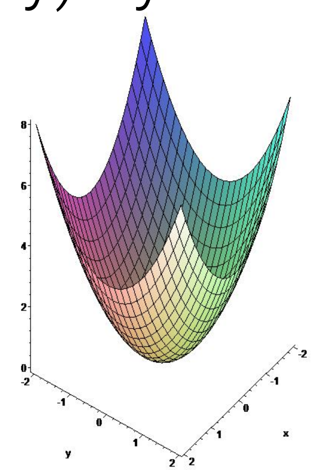
Neg-Log Density

- $\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) + const$
- Quadratic multivariate polynomial



Consequences

- Optimization (maximum probability density) by solving a linear system
- Gaussians are ellipsoids
  - Eigenvectors of Σ are main axes (*principal component analysis*, PCA)
  - Eigenvalues are extremal variances



More Rules for Gaussians

More Rules for Computations with Gaussians

- Products of Gaussians are Gaussians
  - Algorithm: Add quadratic polynomials
  - Variance can only decrease
- Marginals (“projections”) of Gaussians are Gaussians
  - Leave out dimensions in $\boldsymbol{\mu}$, Σ
- Affine mappings of Gaussians are Gaussians
  - Algorithm: apply map to argument $\mathbf{x}$, yields different quadric
- Sums of Gaussians: no closed-form log-densities
- Entropy: $\frac{1}{2}\ln\left((2\pi e)^d \det(\boldsymbol{\Sigma})\right)$

More Rules for Gaussians

Coordinate Transforms

- General Gaussians as affine transforms of unit Gaussians
  - Quadric $\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})+c$
  - Main axis transform:

$$\boldsymbol{\Sigma}^{-1}=\mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{T}}=\mathbf{U}\begin{pmatrix}\sigma_1^{-2} & & \\ & \sigma_2^{-2} & \\ & & \ddots\end{pmatrix}\mathbf{U}^{\mathrm{T}}$$

$$\boldsymbol{\Sigma}^{-\frac{1}{2}}=\mathbf{U}\mathbf{D}^{\frac{1}{2}}\mathbf{U}^{\mathrm{T}}=\mathbf{U}\begin{pmatrix}\sigma_1^{-1} & & \\ & \sigma_1^{-1} & \\ & & \ddots\end{pmatrix}\mathbf{U}^{\mathrm{T}}$$

More Rules for Gaussians

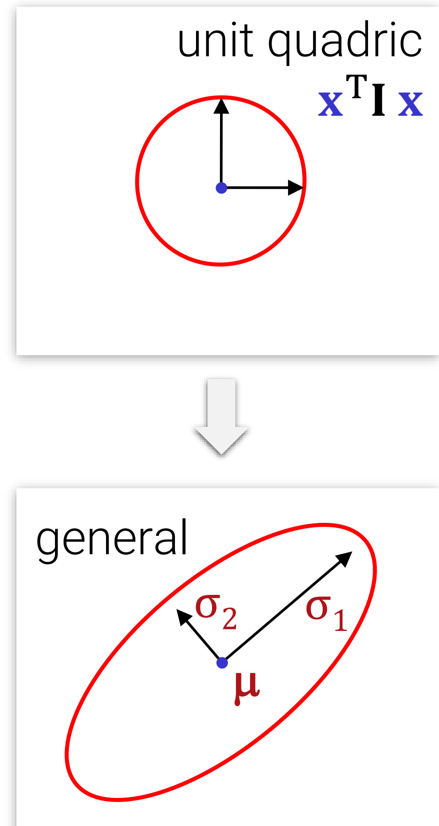
Unit Gaussian

- We get:

$$\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\left(\Sigma^{-\frac{1}{2}}\right)^{\mathrm{T}}\left(\Sigma^{-\frac{1}{2}}\right)(\mathbf{x}-\boldsymbol{\mu})+c$$

$$=\frac{1}{2}\Big(\left(\Sigma^{-\frac{1}{2}}\right)\mathbf{x}-\left(\Sigma^{-\frac{1}{2}}\right)\boldsymbol{\mu}\Big)^{\mathrm{T}}\Big(\left(\Sigma^{-\frac{1}{2}}\right)\mathbf{x}-\left(\Sigma^{-\frac{1}{2}}\right)\boldsymbol{\mu}\Big)+c$$

- This is a unit Quadric / Gaussian $\mathbf{x}^{\mathrm{T}}\mathbf{I}\,\mathbf{x}$
  - rotated to Coordinate frame $\Sigma^{-\frac{1}{2}}$
  - and translated accordingly by $\left(\Sigma^{-\frac{1}{2}}\right)\boldsymbol{\mu}$



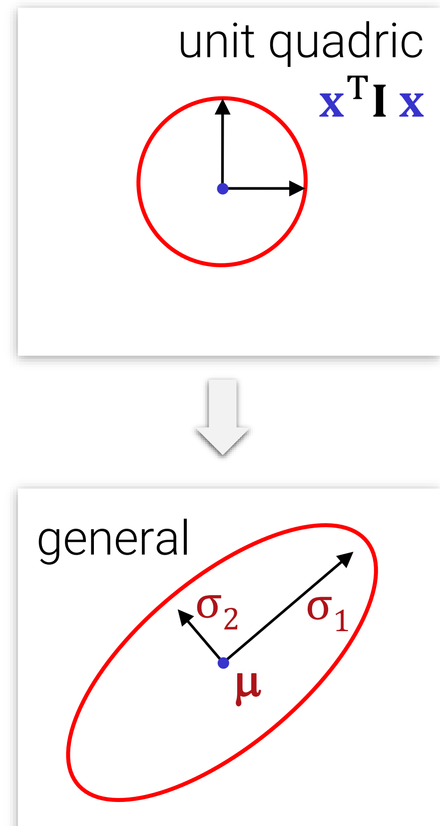
More Rules for Gaussians

Unit Gaussian

- In addition, we have to recompute the (log) normalization factor

$$c = \ln\left(\frac{1}{(2\pi)^{-\frac{d}{2}}\det(\Sigma)^{-\frac{1}{2}}}\right)$$

to ensure a unit integral



Rule of thumb

- All Gaussians are related by
  - Translation
  - Rotation & non-uniform scaling
  - Always adapting the density to integrate to 1

Mahalanobis Distance

Given:

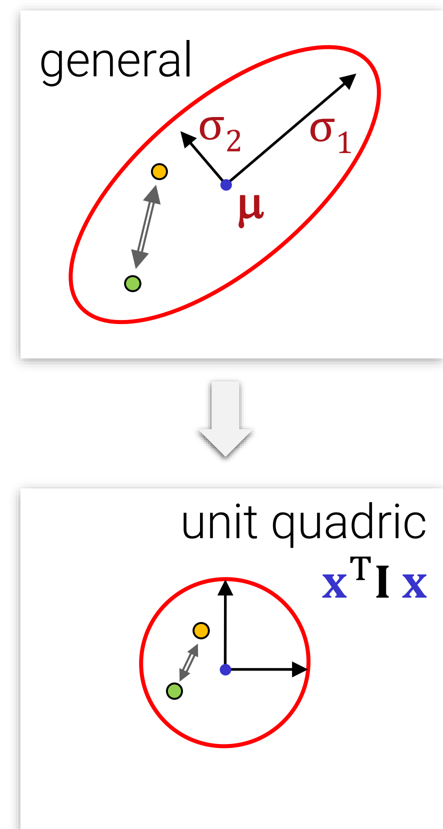
- Gaussian with parameters $\mathbf{\mu}, \mathbf{\Sigma}$
- Sample point $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$

Mahalanobis distance of x

$$D_M(\mathbf{x}) = \sqrt{(\mathbf{x} - \mathbf{\mu})^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \mathbf{\mu})}$$
$$D_M(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \mathbf{y})}$$

Interpretation

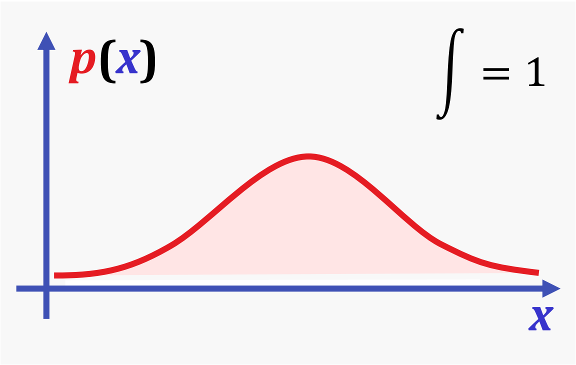
- Distances in “unit Gaussian space”
- One unit = one standard deviation

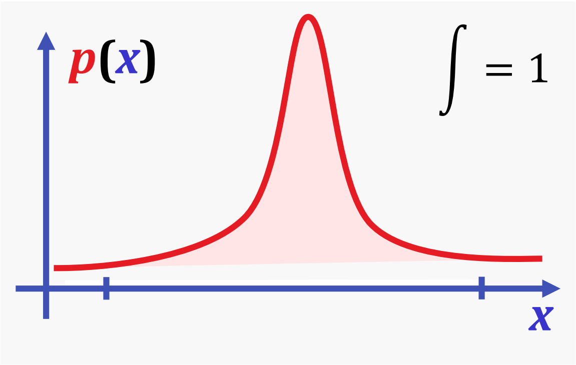


Applications

Example

- Given a sample from and a Gaussian distribution
- How likely is this sample from that distribution?
- Density value not a good measure
  - Absolute density depends on breadth





Estimation from Data

Task

- Data $\mathbf{d}_1, \ldots, \mathbf{d}_\mathbf{n}$ generated w/Gaussian distribution (i.i.d.)
- Estimate parameters

Maximum Likelihood Estimation

- Most likely parameters: $\text{argmax}_{\boldsymbol{\mu},\boldsymbol{\Sigma}} P(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{d}_1, \ldots, \mathbf{d}_n)$

$$\boldsymbol{\mu}_{ml} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{d}_i \qquad \boldsymbol{\Sigma}_{ml} = \frac{1}{n-1}\sum_{i=1}^{n} (\mathbf{d}_i - \boldsymbol{\mu})(\mathbf{d}_i - \boldsymbol{\mu})^{\mathrm{T}}$$

mean covariance