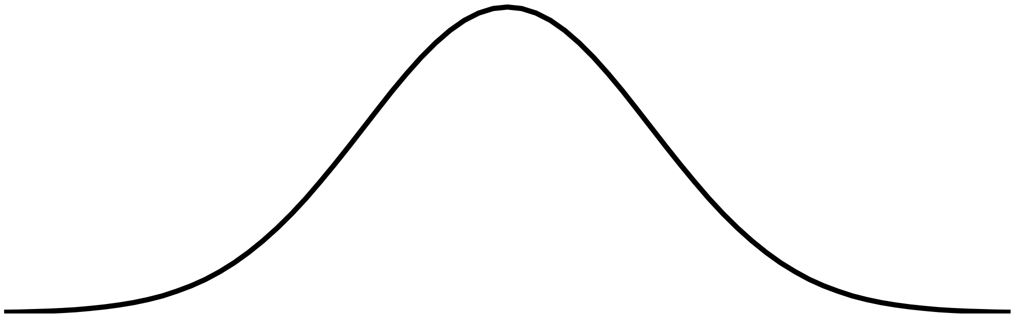
TW414 / PR813 Lecture 3
Gaussians



Ludwig Schwardt

University of Stellenbosch

3 March 2003

1. The Gaussian pdf

- **Multivariate Gaussian (*normal*) pdf defined as**

$$p(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) \triangleq \frac{1}{(2\pi)^{D/2}\left|\boldsymbol{\Sigma}\right|^{1/2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\right\}$$

  evaluated at $D\times 1$ feature vector $\boldsymbol{x}$, with $D\times 1$ *mean vector* $\boldsymbol{\mu}$ and $D \times D$ *covariance matrix* $\boldsymbol{\Sigma}$ satisfying

$$\boldsymbol{\mu} = \mathcal{E}\left[\boldsymbol{x}\right] \quad \text{and} \quad \boldsymbol{\Sigma} = \mathcal{E}\left[(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T\right]$$

- **Also written as $\mathcal{N}(\boldsymbol{\mu},\Sigma)|_{\boldsymbol{x}}$ or $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu},\Sigma)$, where symbol $\sim$ means "has distribution"**

- **Parameters μ and Σ describe center and spread of data**

- **Gaussian pdf has several excellent properties:**

  1. **"Safest bet" amongst all pdfs with given $\boldsymbol{\mu}$ and Σ**
  2. ***Closed under multiplication* (within scale factor):**

$$\mathcal{N}(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1) \times \mathcal{N}(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_2) = C\mathcal{N}(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_1^{-1}\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}\boldsymbol{\Sigma}_2^{-1}\boldsymbol{\mu}_2,\, \boldsymbol{\Sigma}),$$

     with $\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1}$ and $C = \mathcal{N}(\boldsymbol{0},\, \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)|_{\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2}$
  3. ***Closed under convolution* (i.e. sum of independent Gaussian variables is Gaussian):**

$$\mathcal{N}(\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1) * \mathcal{N}(\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_2) = \mathcal{N}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2,\, \boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)$$

  4. **Consequently, for linear systems, $\mathcal{N}$ in $\Longrightarrow \mathcal{N}$ out:**

$$\boldsymbol{A}\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{\mu}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^H) \quad \text{if} \quad \boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$$

  5. **BUT: $\mathcal{N} + \mathcal{N} \neq \mathcal{N}$ (mixture Gaussian instead)**
  6. **Marginal and conditional densities of $\mathcal{N}$ are $\mathcal{N}$**

2. Full covariance Gaussians

- Σ is general matrix $\Longrightarrow$ model has $D + \frac{D(D+1)}{2}$ parameters

Estimate mean μ and covariance Σ from training data as

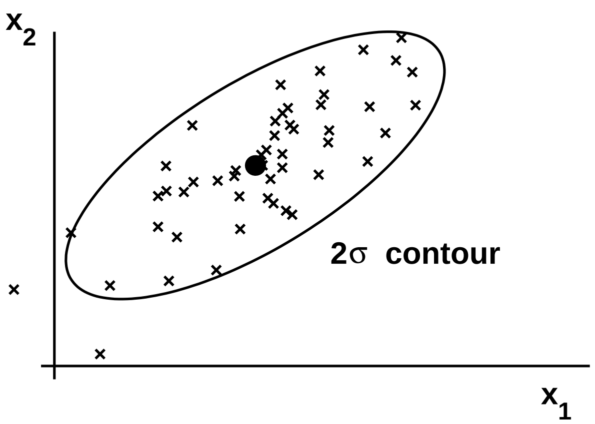
$$\hat{\boldsymbol{\mu}} = \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}_n \quad \textbf{(Matlab } \texttt{mean}\textbf{)}$$

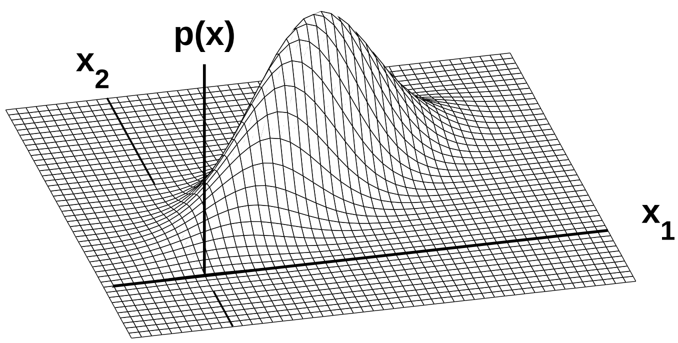
$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N-1}\sum_{n=1}^{N}(\boldsymbol{x}_n - \hat{\boldsymbol{\mu}})(\boldsymbol{x}_n - \hat{\boldsymbol{\mu}})^T \quad \textbf{(Matlab } \texttt{cov}\textbf{)}$$

Score test data with model via *log-likelihood* given by

$$\ln p(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) - \frac{D}{2}\ln(2\pi) - \frac{1}{2}\ln(|\boldsymbol{\Sigma}|)$$

- Models data as ellipsoid-shaped cloud "at an angle" $\Longrightarrow$ allows correlations between feature components
- Most powerful Gaussian model, but needs lots of training data to properly estimate parameters, especially if D high





3. Diagonal covariance Gaussians

- Σ is *diagonal* matrix $\Longrightarrow$ model has $2D$ parameters

Estimate mean μ and $D \times 1$ ***variance vector*** $\boldsymbol{\sigma}^2 = \text{diag}(\Sigma)$ from training data as

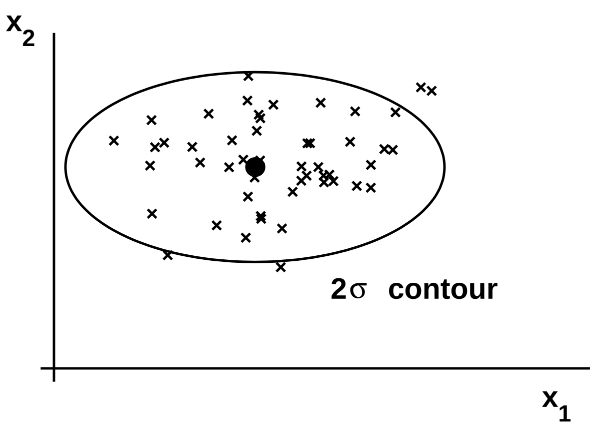
$$\hat{\boldsymbol{\mu}} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n \quad \textbf{(Matlab } \texttt{mean}\textbf{)}$$

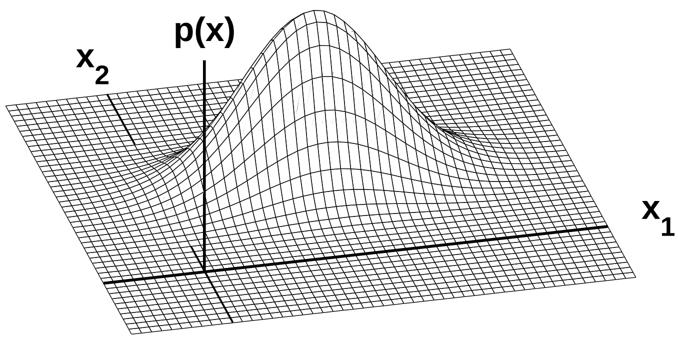
$$\hat{\sigma}_i^2 = \frac{1}{N-1}\sum_{n=1}^{N}(x_{in} - \hat{\mu}_i)^2 \quad \textbf{(Matlab } \texttt{diag(cov(X))}\textbf{)}$$

Score test data with model via ***log-likelihood*** given by

$$\ln p(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\sigma^2}) = -\frac{1}{2}\sum_{i=1}^{D}\frac{(x_i - \mu_i)^2}{\sigma_i^2} - \frac{D}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{D}\ln(\sigma_i^2)$$

- Models data as ellipsoid-shaped cloud aligned with axes $\Longrightarrow$ correlations between feature components ignored
- Good compromise between modelling power and training data hunger, especially when used in mixture models





4. Spherical covariance Gaussians

- $\boldsymbol{\Sigma} = \sigma^2 \boldsymbol{I}$ **(scaled identity)** $\Longrightarrow$ **model has** $D+1$ **parameters**

Estimate mean $\boldsymbol{\mu}$ **and scalar variance** σ^2 **from training data as**

$$\hat{\boldsymbol{\mu}} = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n \quad \textbf{(Matlab } \texttt{mean}\textbf{)}$$

$$\hat{\sigma}^2 = \frac{1}{D(N-1)}\sum_{n=1}^{N} \|\boldsymbol{x}_n - \hat{\boldsymbol{\mu}}\|^2 \quad \textbf{(Matlab } \texttt{mean(diag(cov(X)))}\textbf{)}$$

where $\|\boldsymbol{x}\|^2 \triangleq \boldsymbol{x}^T\boldsymbol{x}$

Score test data with model via *log-likelihood* given by

$$\ln p(\boldsymbol{x}|\boldsymbol{\mu}, \sigma^2) = -\frac{\|\boldsymbol{x}-\boldsymbol{\mu}\|^2}{2\sigma^2} - \frac{D}{2}\ln(2\pi\sigma^2)$$

- **Models data as spherical cloud**
- **Simplest Gaussian pdf** $\Longrightarrow$ **useful when data is very scarce or computational speed and storage is an issue**