Classification

Probabilistic Generative Models II

T.L. Grobler
(with some edits by S. Kroon)

Classification

- Given $\mathbf{x}$ assign to one of k classes:
  - C_j, $j = 1,\ldots,k$
  - Assign prob $P(C_j|\mathbf{x})$
  - $C^* = \text{argmax}_{C_j} P(C_j|\mathbf{x})$
- Class prob, more useful than knowing max class prob.

Generative Approach

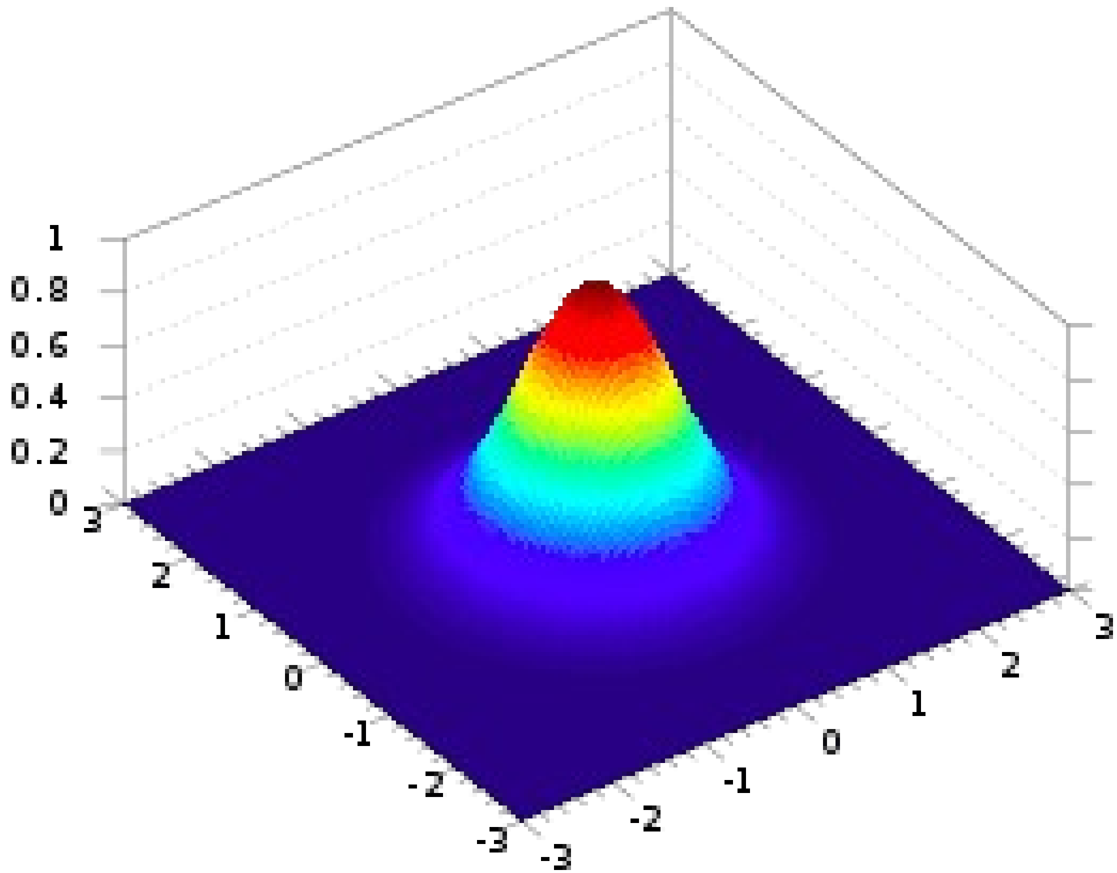
- Estimate class-conditionals: $p(\mathbf{x}|C_j)$
- Posterior (Bayes Theorem):
  - $P(C_j|\mathbf{x}) \propto p(\mathbf{x}|C_j)P(C_j)$
  - Introduce $P(C_j)$
- Probabilistic Generative Models (PGM)

Key steps

- Step 1: **Expand posterior** using logistic/softmax functions.
- Step 2: Investigate/**determine form** of arguments to logistic/softmax functions (under model assumptions).
- Step 3: **estimate parameters** for the resulting form.

Gaussian class-conditional pdfs

$$p(\mathbf{x}|C_j) = \frac{1}{\sqrt{|2\pi\Sigma_j|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{u}_j)^T \Sigma_j^{-1} (\mathbf{x}-\mathbf{u}_j) \right)$$



Two Classes: Shared Σ

$$P(C_1|\mathbf{x}) = \sigma(\mathbf{w}^T\mathbf{x} + w_0)$$

$$a(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + w_0$$

$$\mathbf{w} = \Sigma^{-1}(\mathbf{u}_1 - \mathbf{u}_2)$$

$$w_0 = -\frac{1}{2}\mathbf{u}_1^T\Sigma^{-1}\mathbf{u}_1 + \frac{1}{2}\mathbf{u}_2^T\Sigma^{-1}\mathbf{u}_2 + \ln\frac{P(C_1)}{P(C_2)}$$

Linear Decision Boundary

$$
\begin{aligned}
P(C_1|\mathbf{x}) &= P(C_2|\mathbf{x}) = 1 - P(C_1|\mathbf{x}) \\
\sigma(\mathbf{w}^T\mathbf{x} + w_0) &= 1 - \sigma(\mathbf{w}^T\mathbf{x} + w_0) \\
&= \frac{1}{2} \\
\mathbf{w}^T\mathbf{x} + w_0 &= 0
\end{aligned}
$$

Parameter Estimation

- Estimate:
  - Parameters of class-conditional densities
  - Given observations
  - Maximum Likelihood Estimation (MLE)

$$\pi = \frac{N_1}{N} \qquad \mathbf{u}_1 = \frac{1}{N_1}\sum_{n=1}^{N} y_n \mathbf{x}_n \qquad \mathbf{u}_2 = \frac{1}{N_2}\sum_{n=1}^{N}(1-y_n)\mathbf{x}_n$$

$$\Sigma_1 = \tfrac{1}{N_1}\Sigma_{n\in C_1}(\mathbf{x}_n - \mathbf{u}_1)(\mathbf{x}_n - \mathbf{u}_1)^T$$

$$\Sigma_2 = \tfrac{1}{N_2}\Sigma_{n\in C_2}(\mathbf{x}_n - \mathbf{u}_2)(\mathbf{x}_n - \mathbf{u}_2)^T$$

Introduction: Naive Bayes

- MLE Expensive
  - d dim in k classes
  - $1/2kd(d+3)$ – $\mathbf{u}$, Σ
- Share Σ
  - $kd+1/2d(d+1)$
- Diagonal Σ
  - Naive Bayes
  - $2kd$

Conditional Independence

- Features $\mathbf{x}$
  - d dim
  - Conditionally independent
  - Diagonal $\boldsymbol{\Sigma}$

$$p(\mathbf{x}|C) = \prod_{n=1}^{d} p(x_n|C)$$

Diagonal Σ

$$P(\mathbf{x}|C) = \prod_n P(x_n|C)$$

$$= \prod_n \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{1}{2}\frac{(x_n - u_n)^2}{\sigma_n^2}\right)$$

$$= \frac{1}{\sqrt{|2\pi\Sigma|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{u})^T\Sigma^{-1}(\mathbf{x}-\mathbf{u})\right)$$

Diagonal

Naive Bayes Derivation

$$P(C_j|\mathbf{x}) = \frac{P(C_j)p(\mathbf{x}|C_j)}{p(x)}$$

$$= \frac{P(C_j)\prod_n p(x_n|C_j)}{\sum_i P(C_i)\prod_n p(x_n|C_i)}$$

$$\boxed{C^* = \operatorname{argmax}_{C_j} P(C_j)\prod_n p(x_n|C_j)}$$

Naive Bayes Parameter Estimates

$$P(C_j) = \frac{N_j}{N}$$

$$u_{nj} = \frac{1}{N_j} \sum x_{nj}$$

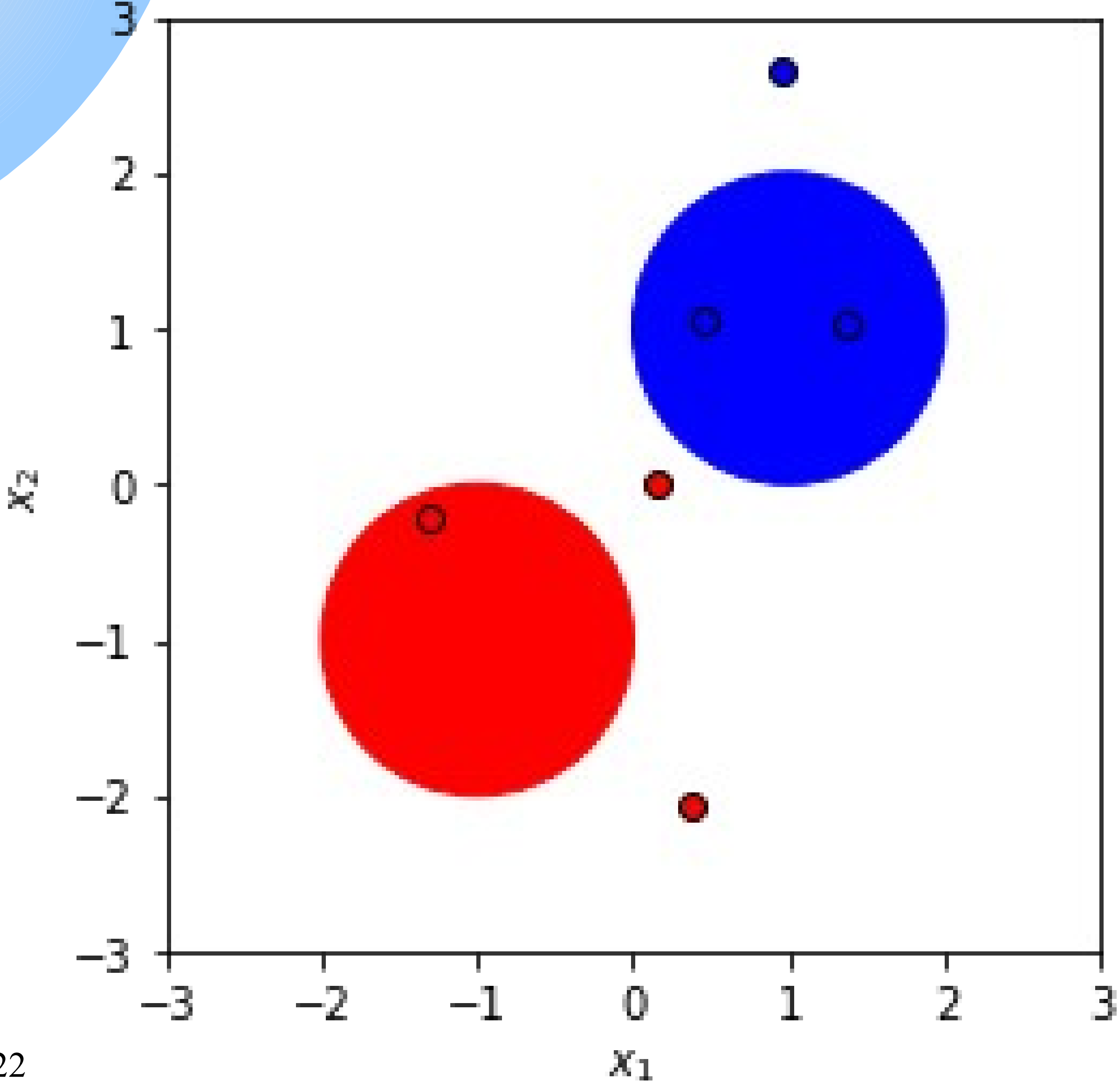
$$\sigma_{nj}^2 = \frac{1}{N_j} \sum (x_{nj} - u_{nj})^2$$

Example

	C_1	C_2
$\mathbf{x}_1$	$[0.3682,-2.0530]^T$	$[0.9456,26543]^T$
$\mathbf{x}_2$	$[0.1521,0.0131]^T$	$[1.3574,1.0225]^T$
$\mathbf{x}_3$	$[-1.3033,-0.2105]^T$	$[0.4478,1.0543]^T$

- Data
  - 2 Classes, 2 Features
  - Gaussian class-conditional pdfs
    - **u** (-1,-1) and (1,1)
    - Same σ, σ=1

Data



Mean and Var

	C_1	C_2
u	$[-0.2610,-0.7501]^T$	$[0.9169, 1.5770]^T$
σ^2	0.7291^2	...

$$\mathbf{u}_j^n = \frac{1}{t}\sum_{s=1}^{t}\mathbf{x}_{sj}^n \qquad \sigma^2 = \frac{1}{kdt}\sum_{j=1}^{k}\sum_{n=1}^{d}\sum_{s=1}^{t}(\mathbf{x}_{sj}^n - \mathbf{u}_j^n)^2$$

- j – class index
- n – feature dimension index
- s – sample index

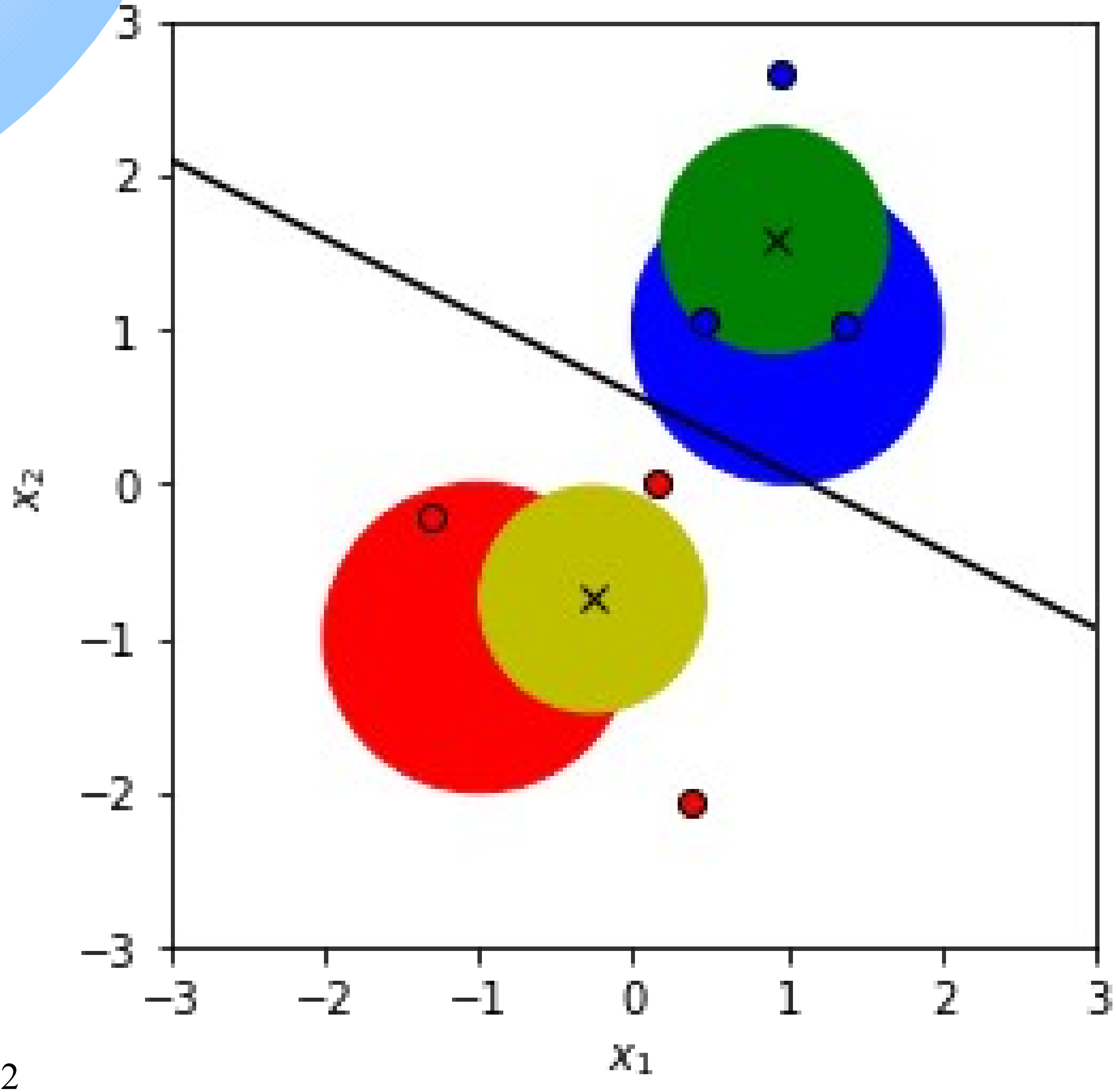
Weights

Parameter	Value
w	[-2.2154,-4.377]
w_0	2.5362
m	-0.506
c	0.5795

$$\mathbf{w} = \frac{1}{\sigma^2}(\mathbf{u}_1 - \mathbf{u}_2) \qquad w_0 = -\frac{1}{2\sigma^2}(\mathbf{u}_1^T\mathbf{u}_1 - \mathbf{u}_2^T\mathbf{u}_2)$$

$$\begin{aligned} x_2 &= \frac{-w_1}{w_2}x_1 - \frac{w_0}{w_2} \\ &= mx_1 + c \end{aligned}$$

Decision Boundary



Maximum Likelihood

$$
\begin{aligned}
P(\mathbf{x}|\theta) &= \prod_s p(\mathbf{x}_s|y_s)P(y_s) \\
&= \prod_s \left(\prod_n p(x_{ns}|y_s) \right) P(y_s) \\
&= \prod_s \left(\prod_n \left(\prod_j p(x_{ns}|C_j)^{\mathbf{1}_{y_s=C_j}} \right) \right) \prod_j \left(P(C_j)^{\mathbf{1}} \right)
\end{aligned}
$$

$$
\mathbf{1}_{y_s=C_j} = \begin{cases} 1 & y_s = C_j \\ 0 & \text{otherwise} \end{cases}
$$

Indicator Function

Log-likelihood

$$\log P(\mathbf{x}|\theta) = f_1(\mathbf{x},\theta) + f_2(\mathbf{x},\theta)$$

$$f_1(\mathbf{x},\theta) = \sum_s \sum_n \sum_j \mathbf{1} \log(P(x_{ns}|C_j))$$

$$f_2(\mathbf{x},\theta) = \sum_s \sum_j \mathbf{1} \log(P(C_j))$$

$$P(x_{ns}|C_j) = \frac{1}{\sigma_{nj}\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x_{ns}-u_{nj}}{\sigma_{nj}}\right)^2\right)$$

$$\log(P(x_{ns}|C_j)) = -\log\left(\sqrt{2\pi}\right) - \log(\sigma_{nj}) - \frac{1}{2}\left(\frac{x_{ns}-u_{nj}}{\sigma_{nj}}\right)^2$$

Mean

$$\frac{\partial f_1(\mathbf{x},\theta)}{\partial u_{nj}} = \sum_s -\mathbf{1}\frac{(x_{ns}-u_{nj})}{\sigma_{nj}^2}$$
$$= 0$$
$$\widehat{u}_{nj} = \frac{1}{N_j}\sum_s \mathbf{1} x_{ns}$$

Variance

$$\frac{\partial f_1(\mathbf{x}, \theta)}{\partial \sigma_{nj}} = \sum_s -\frac{\mathbf{1}}{\sigma_{nj}} + \mathbf{1}\frac{(x_{ns} - u_{nj})^2}{\sigma_{nj}^3}$$

$$= 0$$

$$\widehat{\sigma}_{nj}^2 = \frac{1}{N_j}\sum_s \mathbf{1}(x_{ns} - \widehat{u}_{nj})^2$$

Prior

$$f_3(\mathbf{x},\theta) = \sum_s \sum_j \mathbf{1} \log(P(C_j)) + \lambda \left(\sum_j P(C_j) - 1 \right)$$

Lagrange

$$\frac{\partial f_3(\mathbf{x},\theta)}{\partial P(C_j)} = \sum_s \frac{\mathbf{1}}{P(C_j)} + \lambda = 0$$

$$P(C_j) = -\frac{\sum_s \mathbf{1}}{\lambda}$$

$$\sum_j P(C_j) = -\frac{\sum_s \sum_j \mathbf{1}}{\lambda}$$

Constraint: 1

$$\lambda = -N$$

$$\widehat{P}(C_j) = \frac{N_j}{N}$$