Aprendizagem Automática

João Paulo Pordeus Gomes

Métodos Estatísticos

Aula Anterior

- Histogramas e distribuições
- Distribuição Gaussiana
- Regra de Bayes

Classificadores Estatísticos

Classificação

- Observar uma série de características de uma instância e determinar a classe a qual este pertence

Classificação

- Observar uma série de características de uma instância e determinar a classe a qual este pertence

Classificação

- Observar uma série de características (x) de uma instância e determinar a classe (c) a qual este pertence

Classificação

- Observar uma série de características (x) de uma instância e determinar a classe (c) a qual este pertence
- p(c|x)

Classificação

- Observar uma série de características (x) de uma instância e determinar a classe (c) a qual este pertence
- $p(c|x) = p(x|c)p(c) / p(x)$

Classificação

- Caso com duas classes (c_1 e c_2)

Classificação

- Caso com duas classes (c_1 e c_2)
  - $p(c_1|x) = p(x|c_1)p(c_1) / p(x)$
  - $p(c_2|x) = p(x|c_2)p(c_2) / p(x)$

Classificação

- Caso com duas classes (c_1 e c_2)
  - $p(c_1|x) = p(x|c_1)p(c_1) / p(x)$
  - $p(c_2|x) = p(x|c_2)p(c_2) / p(x)$
- Para classificar, escolhe o mais provável

Classificação

- Caso com duas classes (c_1 e c_2)
  - $p(c_1|x) = p(x|c_1)p(c_1) / p(x)$
  - $p(c_2|x) = p(x|c_2)p(c_2) / p(x)$

Classificação

- Caso com duas classes (c_1 e c_2)
  - $p(c_1|x) \propto p(x|c_1)p(c_1)$
  - $p(c_2|x) \propto p(x|c_2)p(c_2)$

Classificação

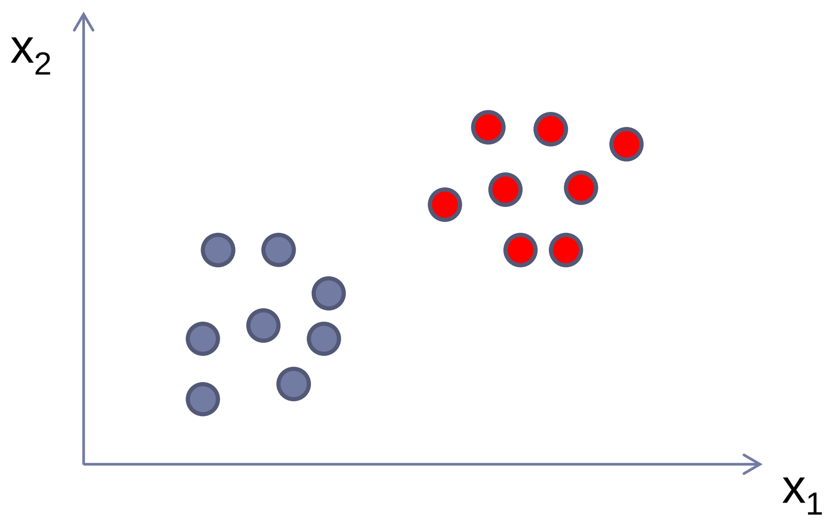
- Caso com duas classes (c_1 e c_2)
  - $p(c_1|x) \propto p(x|c_1)p(c_1)$
  - $p(c_2|x) \propto p(x|c_2)p(c_2)$

- Estimar $p(c_2)$, $p(c_1)$, $p(x|c_2)$ e $p(x|c_1)$ a partir do conjunto de treino

Classificação

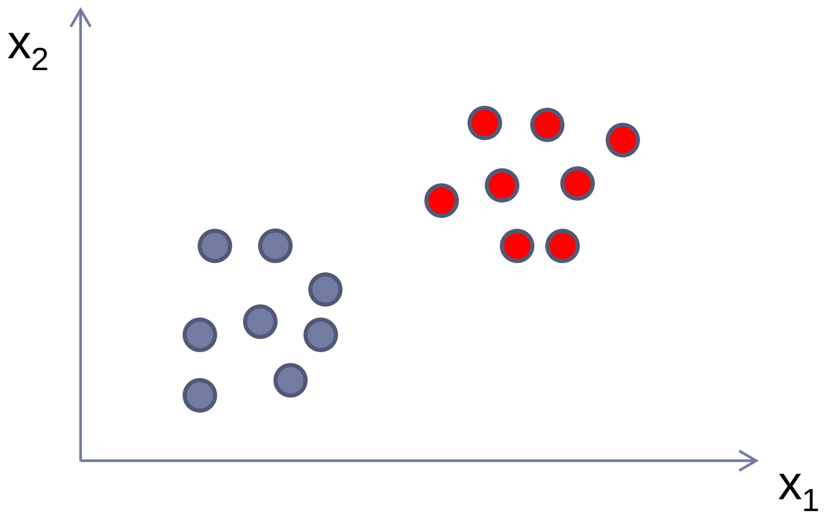
- Caso com duas classes (c_1 e c_2)
  - $p(c_1|x) \propto p(x|c_1)p(c_1)$
  - $p(c_2|x) \propto p(x|c_2)p(c_2)$

- Estimar $p(c_2)$, $p(c_1)$, $p(x|c_2)$ e $p(x|c_1)$ a partir do conjunto de treino



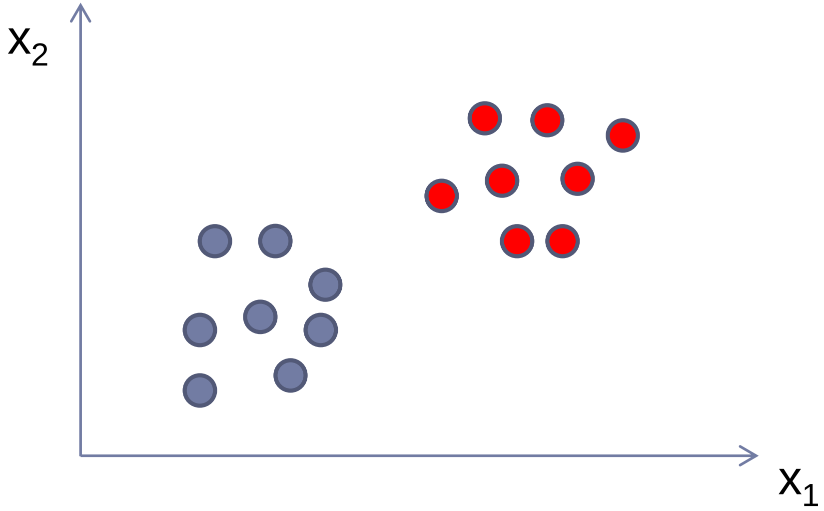
Classificação

- $p(c_2)$ e $p(c_1)$



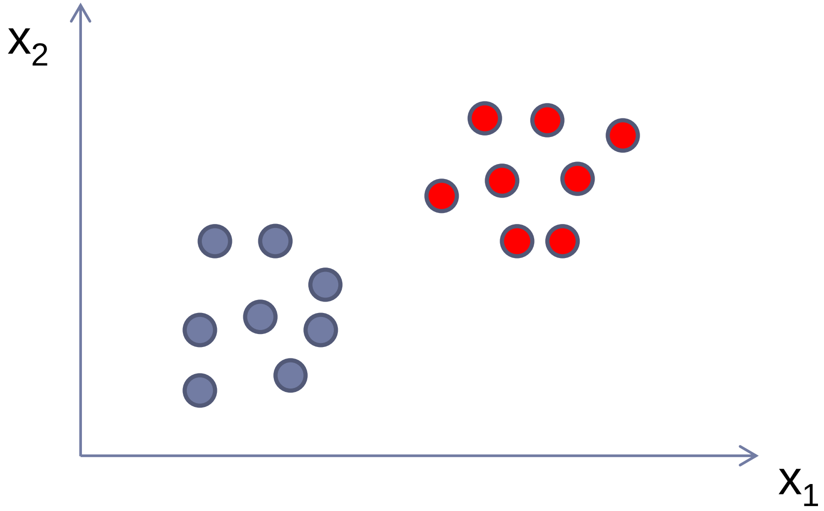
Classificação

- $p(c_2)$ e $p(c_1)$
  - Classes equiprováveis
  - Proporcional ao número de exemplos
  - Conhecido



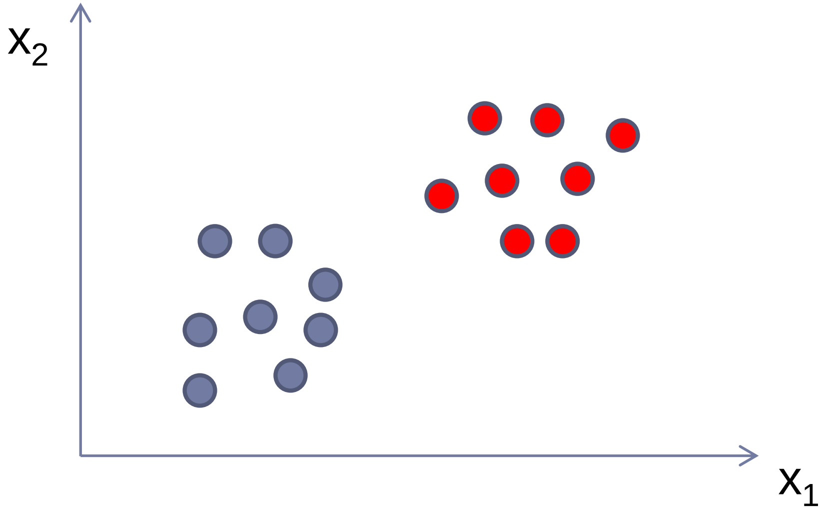
Classificação

- $p(x|c_2)$ e $p(x|c_1)$



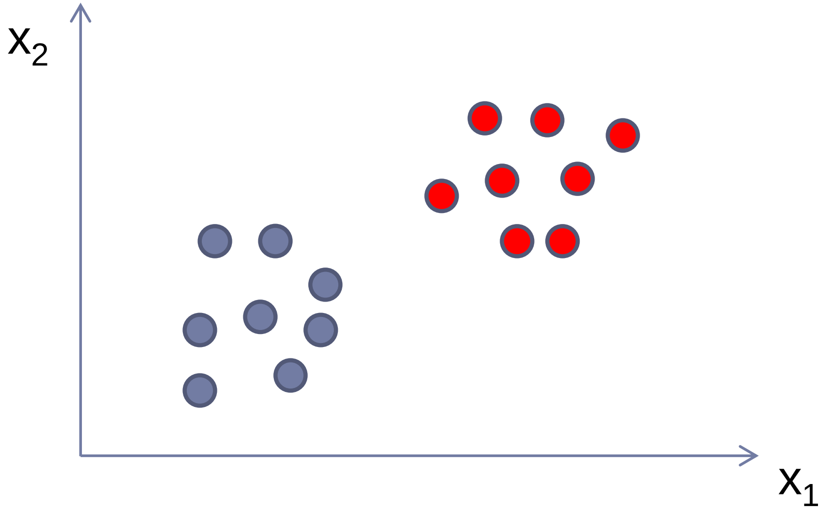
Classificação

- $p(x|c_2)$ e $p(x|c_1)$
  - Supõe que os dados são Gaussianos



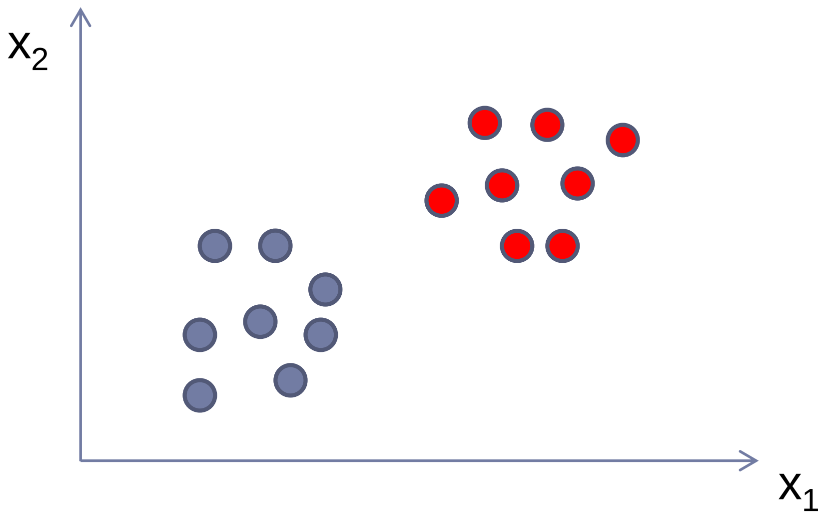
Classificação

- $p(x|c_2)$ e $p(x|c_1)$
  - Supõe que os dados são Gaussianos
  - Estima os parâmetros da distribuição



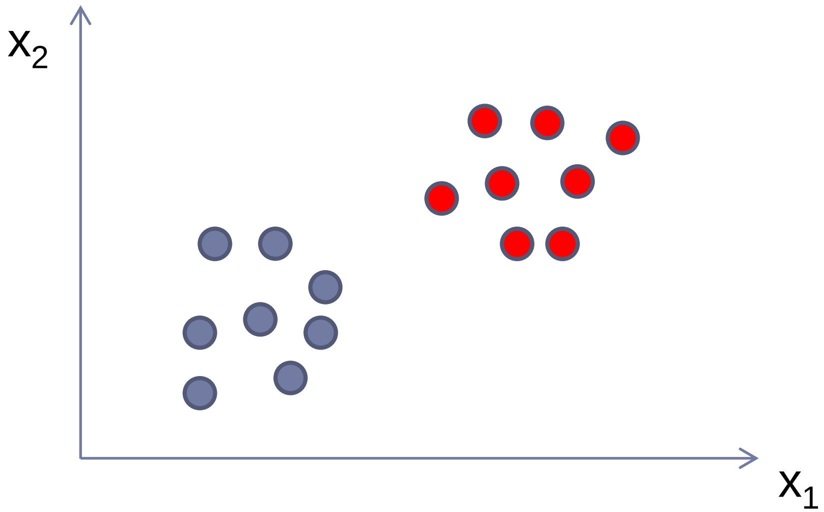
Classificação

- $p(x|c_2)$ e $p(x|c_1)$
  - Supõe que os dados são Gaussianos
  - Estima os parâmetros da distribuição
  - Calcula $p(x|c_2)$ e $p(x|c_1)$



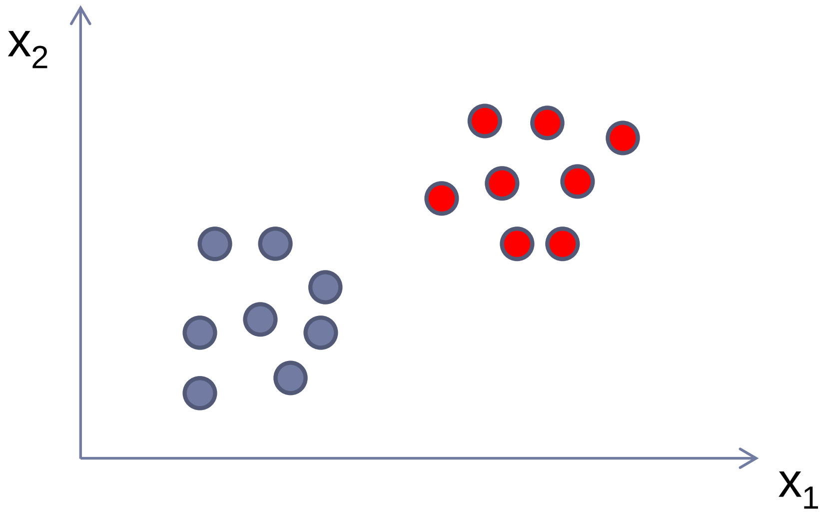
Classificação

- Para cada classe
  - $p(c) = n_c / N$ ou $p(c) = 0.5$



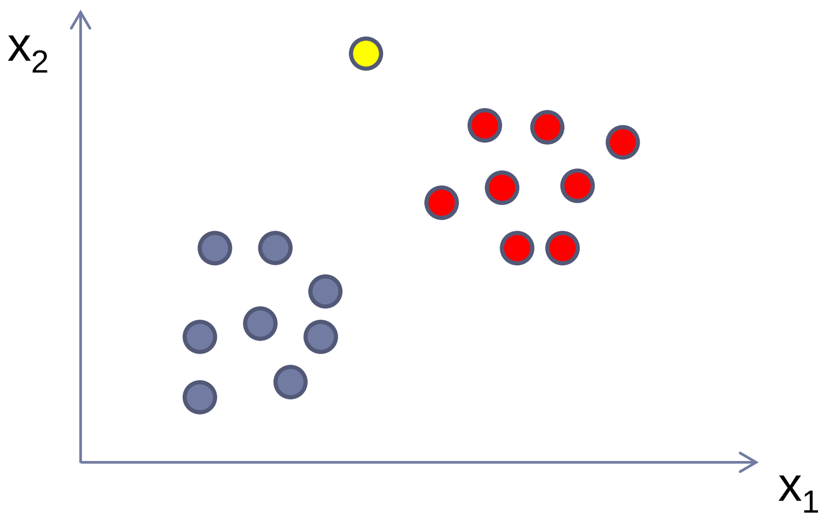
Classificação

- Para cada classe
  - p(c) = n_c / N ou p(c) = 0.5
  - Calcula
    - $\boldsymbol{\mu}_{\boldsymbol{c}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i$ e $\Sigma_c = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T$



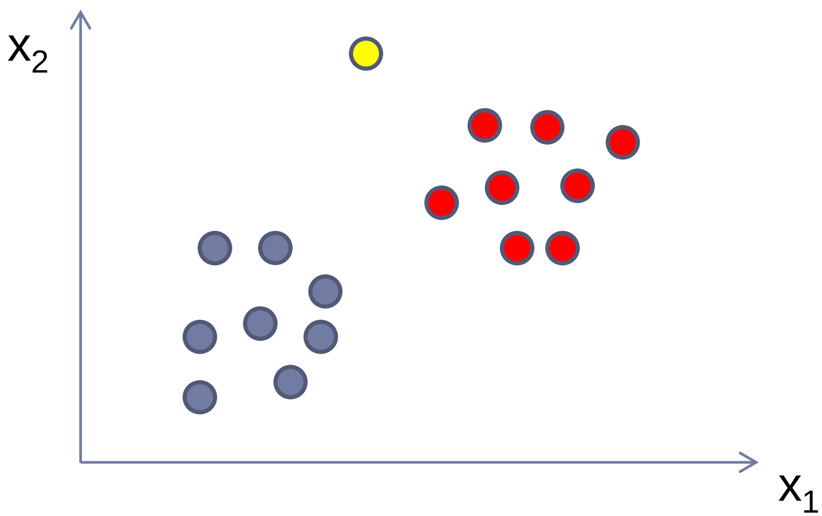
Classificação

- Quando se deseja classificar



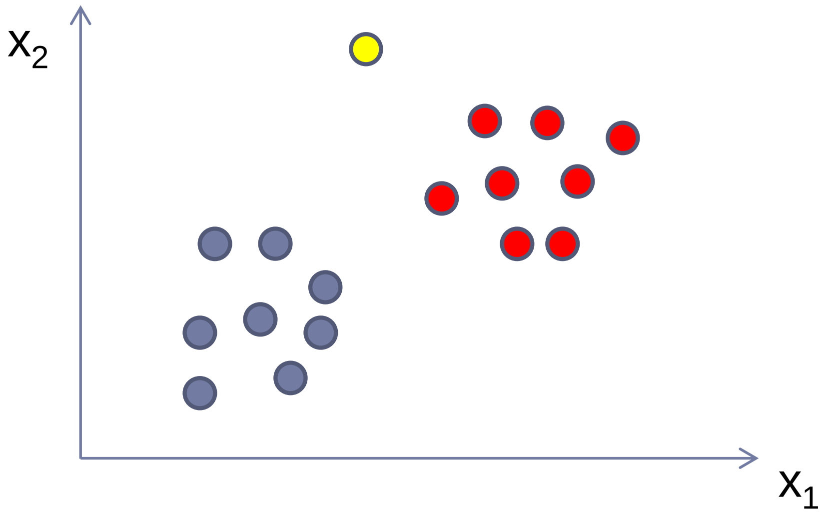
Classificação

- Quando se deseja classificar
- Para cada classe
  - $p(\boldsymbol{x}|c) = \frac{1}{|\Sigma|^{1/2}(2\pi)^{m/2}} exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu_c})^T \Sigma^{-1}{}_c (\boldsymbol{x}-\boldsymbol{\mu_c})\right)$



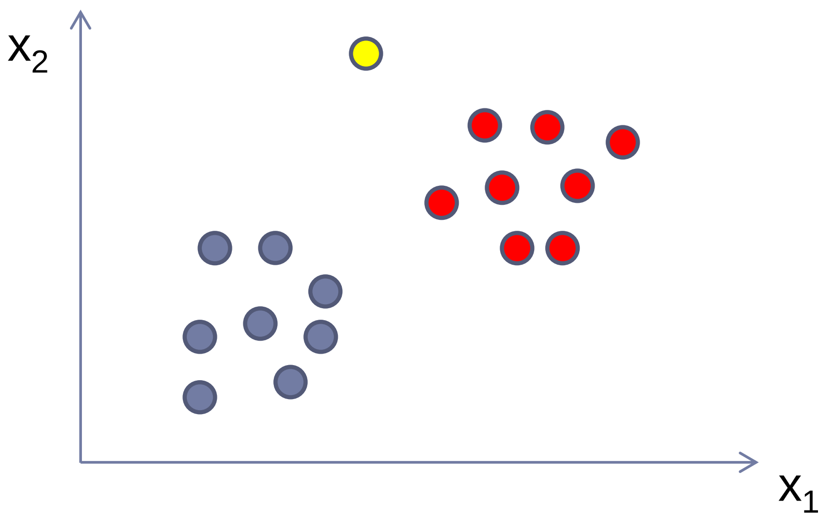
Classificação

- Quando se deseja classificar
- Para cada classe
  - $p(\boldsymbol{x}|c) = \frac{1}{|\Sigma|^{1/2}(2\pi)^{m/2}} exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu_c})^T {\Sigma^{-1}}_c(\boldsymbol{x}-\boldsymbol{\mu_c})\right)$
  - $p(c|x) \propto p(x|c)p(c)$



Classificação

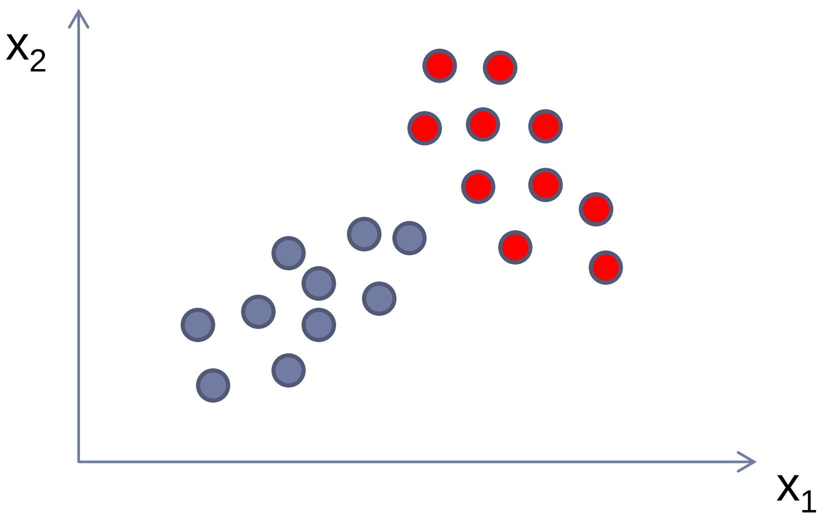
- Quando se deseja classificar
- Para cada classe
  - $p(\boldsymbol{x}|c) = \frac{1}{|\Sigma|^{1/2}(2\pi)^{m/2}} exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu_c})^T {\Sigma^{-1}}_c (\boldsymbol{x}-\boldsymbol{\mu_c})\right)$
  - $p(c|x) \propto p(x|c)p(c)$
  - Escolhe o mais provável



Análise do Classificador Gaussiano

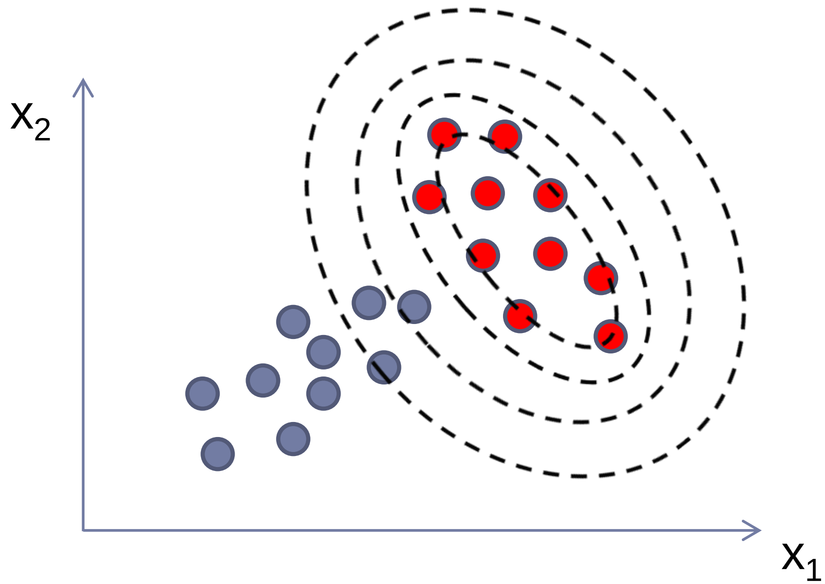
Classificação

- $p(c|x) \propto p(x|c)p(c)$



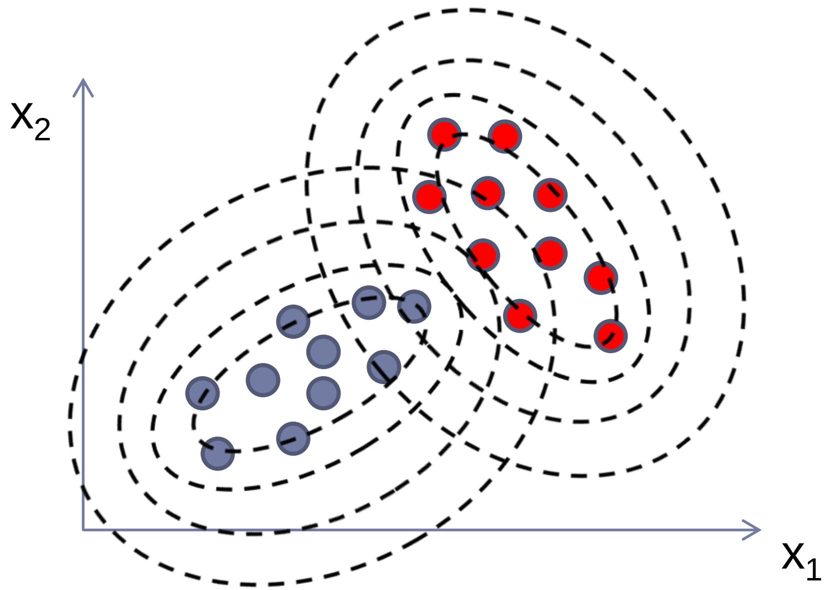
Classificação

- $p(c|x) \propto p(x|c)p(c)$



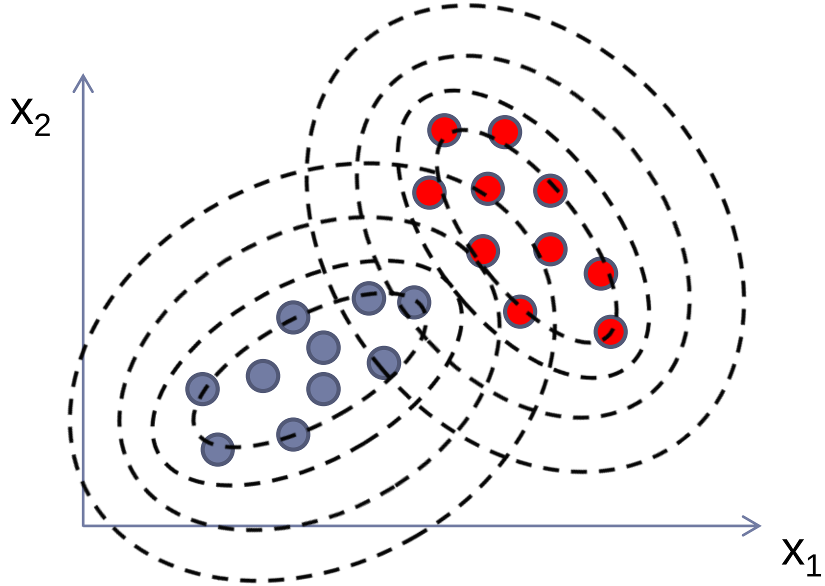
Classificação

- $p(c|x) \propto p(x|c)p(c)$



Classificação

- $p(c|x) \propto p(x|c)p(c)$



Discriminante Quadrático Gaussiano

Discriminante Quadrático Gaussiano

- p(c|x) ∝ p(x|c)p(c)
- Dados distribuídos segundo uma gaussiana
  - $\boldsymbol{\mu}_{\boldsymbol{c}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i$
  - $\Sigma_c = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T$

Matriz de Covariância

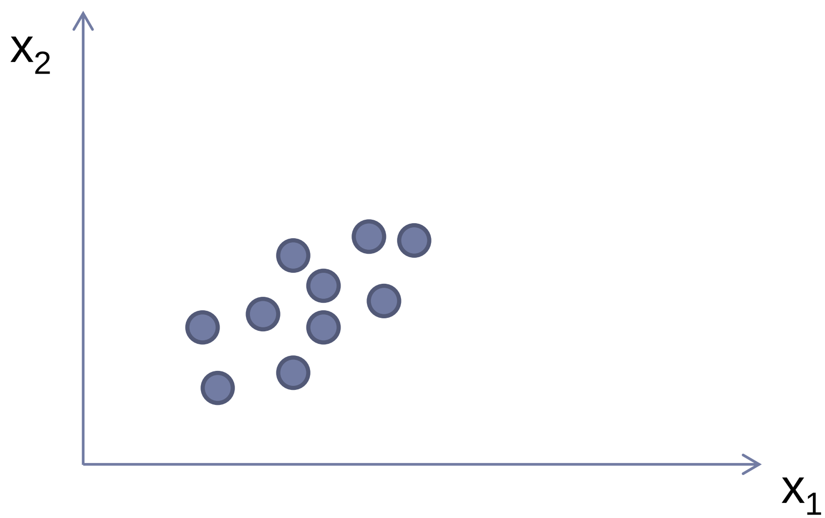
- $\Sigma_c = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T$
- Para um vetor de dimensão 2 a matriz é 2x2
  - $\Sigma = \begin{bmatrix} \sigma_1{}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2{}^2 \end{bmatrix}$

Matriz de Covariância

- $\boldsymbol{\Sigma}_c = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T$
- Para um vetor de dimensão 2 a matriz é 2x2
  - $\Sigma = \begin{bmatrix} \sigma_1{}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2{}^2 \end{bmatrix}$
    - Valores altos de σ_{12}

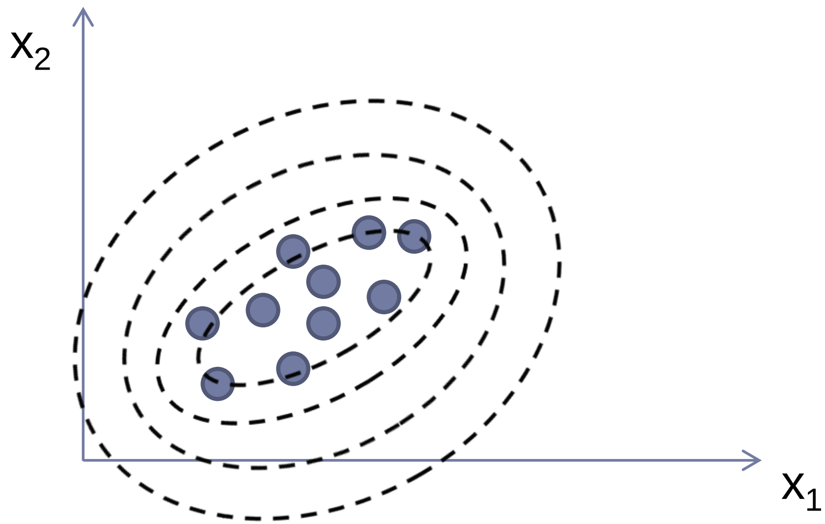
Matriz de Covariância

- $\Sigma_c = \frac{1}{n-1} \sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T$
- Para um vetor de dimensão 2 a matriz é 2x2
  - $\Sigma = \begin{bmatrix} \sigma_1{}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2{}^2 \end{bmatrix}$
    - Valores altos de σ_{12}



Matriz de Covariância

- $\Sigma_c = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T$
- Para um vetor de dimensão 2 a matriz é 2x2
  - $\Sigma = \begin{bmatrix} \sigma_1{}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2{}^2 \end{bmatrix}$
    - Valores altos de σ_{12}

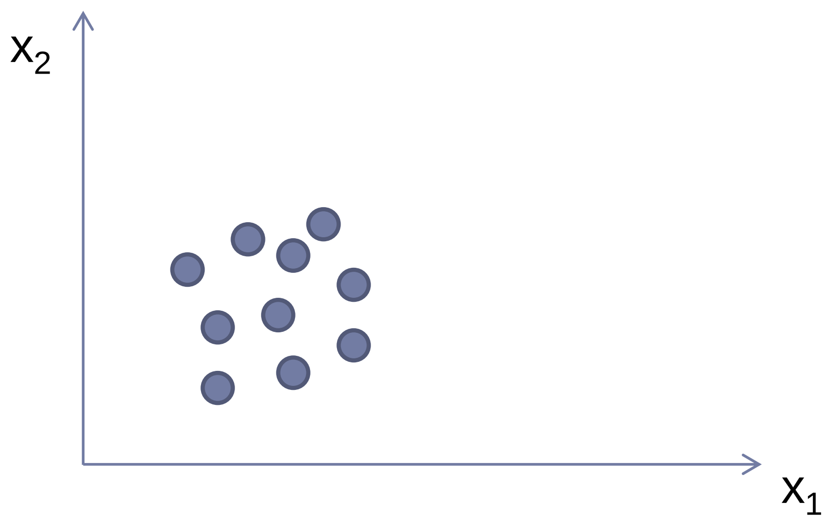


Matriz de Covariância

- $\Sigma_c = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T$
- Para um vetor de dimensão 2 a matriz é 2x2
  - $\Sigma = \begin{bmatrix} \sigma_1{}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2{}^2 \end{bmatrix}$
    - Valores de σ_{12} próximos de 0

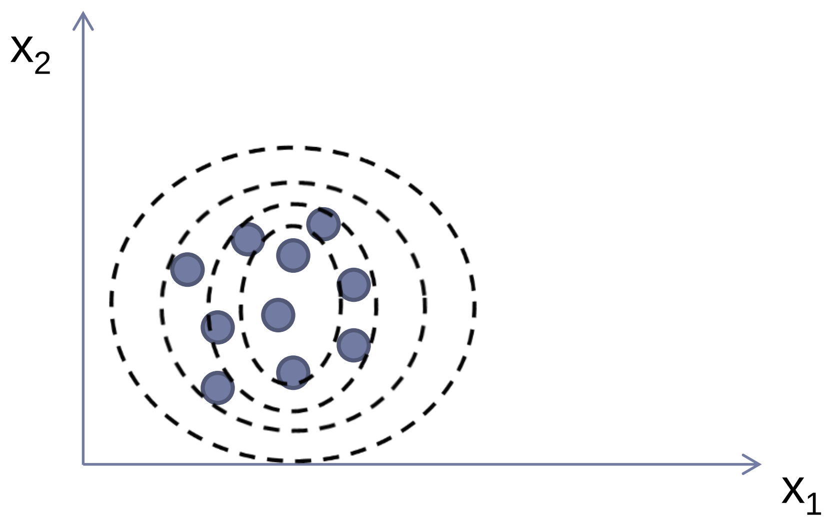
Matriz de Covariância

- $\Sigma_c = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T$
- Para um vetor de dimensão 2 a matriz é 2x2
  - $\Sigma = \begin{bmatrix} \sigma_1{}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2{}^2 \end{bmatrix}$
    - Valores de σ_{12} próximos de 0



Matriz de Covariância

- $\Sigma_c = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})^T$
- Para um vetor de dimensão 2 a matriz é 2x2
  - $\Sigma = \begin{bmatrix} \sigma_1{}^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2{}^2 \end{bmatrix}$
    - Valores de σ_{12} próximos de 0

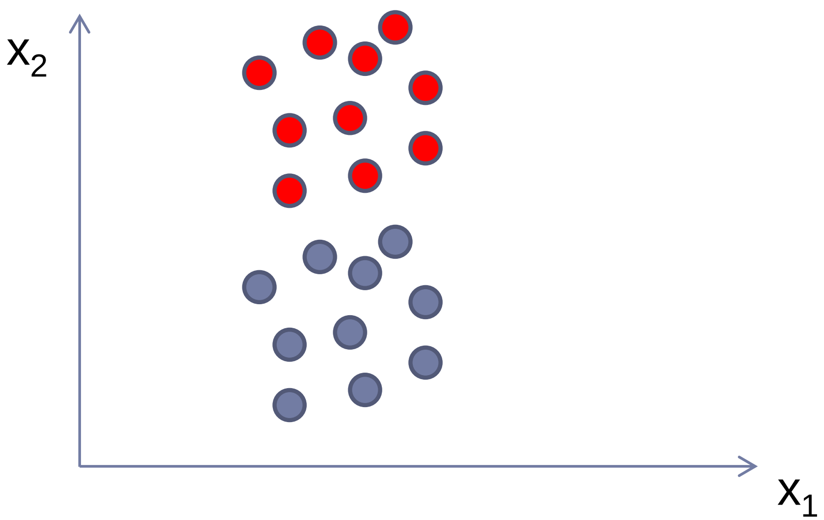


Naive Bayes Gaussiano

- Considera que as variáveis não tem correlação
- Matriz de covariância diagonal
  - $\Sigma = \begin{bmatrix} {\sigma_1}^2 & 0 \\ 0 & {\sigma_2}^2 \end{bmatrix}$

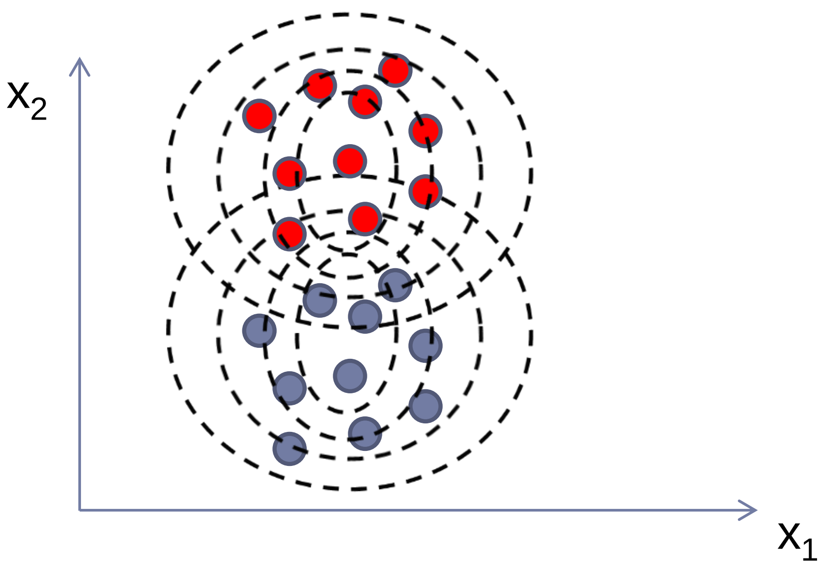
Naive Bayes Gaussiano

- Considera que as variáveis não tem correlação
- Matriz de covariância diagonal
  - $\Sigma = \begin{bmatrix} {\sigma_1}^2 & 0 \\ 0 & {\sigma_2}^2 \end{bmatrix}$



Naive Bayes Gaussiano

- Considera que as variáveis não tem correlação
- Matriz de covariância diagonal
  - $\Sigma = \begin{bmatrix} {\sigma_1}^2 & 0 \\ 0 & {\sigma_2}^2 \end{bmatrix}$



Naive Bayes – Formulação Geral

Naive Bayes

- Formulação Geral
  - $p(c|\boldsymbol{x}) \propto p(\boldsymbol{x}|c)p(c)$

Naive Bayes

- Formulação Geral
  - $p(c|\boldsymbol{x}) \propto p(\boldsymbol{x}|c)p(c)$
  - $\boldsymbol{x} = [x_1 \quad \cdots \quad x_n]$

Naive Bayes

- Formulação Geral
  - $p(c|\boldsymbol{x}) \propto p(\boldsymbol{x}|c)p(c)$
  - $\boldsymbol{x} = [x_1 \quad \cdots \quad x_n]$
- Como $x_1 \perp x_2 \perp \ldots \perp x_n$ .Podemos reescrever
  - $p(c|\boldsymbol{x}) \propto p(x_1|c)p(x_2|c) \ldots p(x_n|c)p(c)$

Naive Bayes

- Formulação Geral
  - $p(c|\boldsymbol{x}) \propto p(\boldsymbol{x}|c)p(c)$
  - $\boldsymbol{x} = [x_1 \quad \dots \quad x_n]$
- Como $x_1 \perp x_2 \perp \dots \perp x_n$ .Podemos reescrever
  - $p(c|\boldsymbol{x}) \propto p(x_1|c)p(x_2|c)\dots p(x_n|c)p(c)$
  - $p(c|\boldsymbol{x}) \propto p(c)\prod_{i=1}^{n} p(x_i|c)$

Dúvidas ?