



Aprendizagem Automática



João Paulo Pordeus Gomes

A vertical blue bar is located on the left side of the slide, partially enclosed by a thin white rectangular border.

Aula Anterior

Regressão Logística

Regressão Logística

- ▶ Modelo

- ▶ $\bar{y}_i = \frac{1}{1+e^{-\mathbf{w}^T \mathbf{x}_i}}$

- ▶ Regra de Aprendizado

- ▶ Gradiente descendente

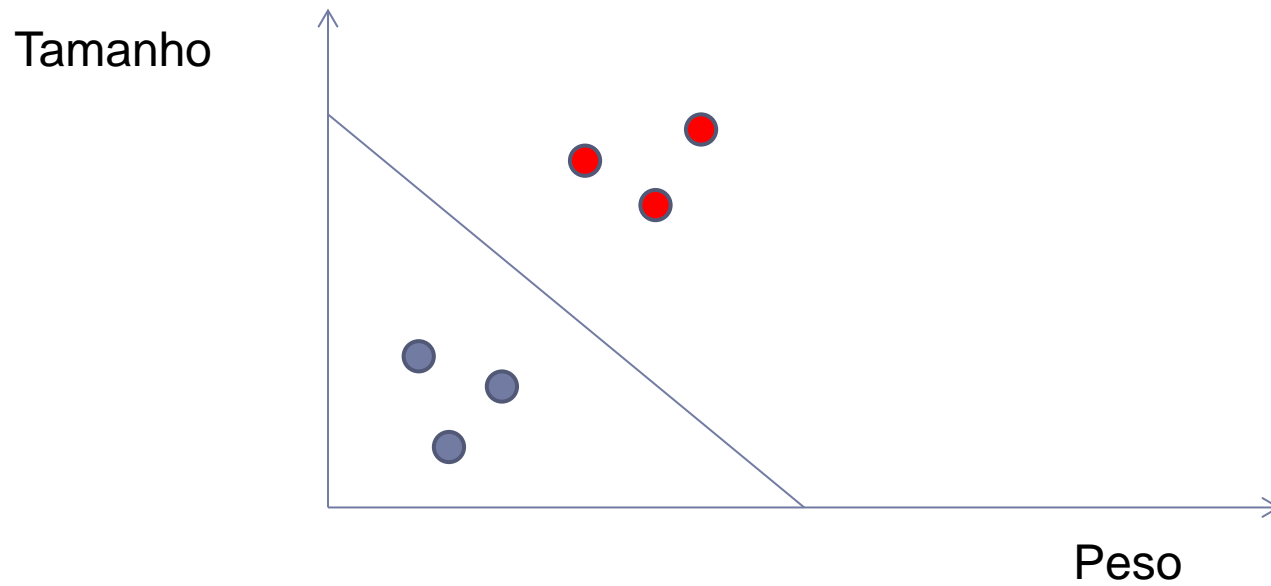
- ▶ $\mathbf{w} = \mathbf{w} + \alpha \frac{1}{n} \sum_{i=1}^n e_i \mathbf{x}_i$

- ▶ Gradiente descendente estocástico

- ▶ $\mathbf{w} = \mathbf{w} + \alpha e_i \mathbf{x}_i$



Regressão Logística



A vertical blue bar is located on the left side of the slide, partially enclosed by a thin white rectangular border.

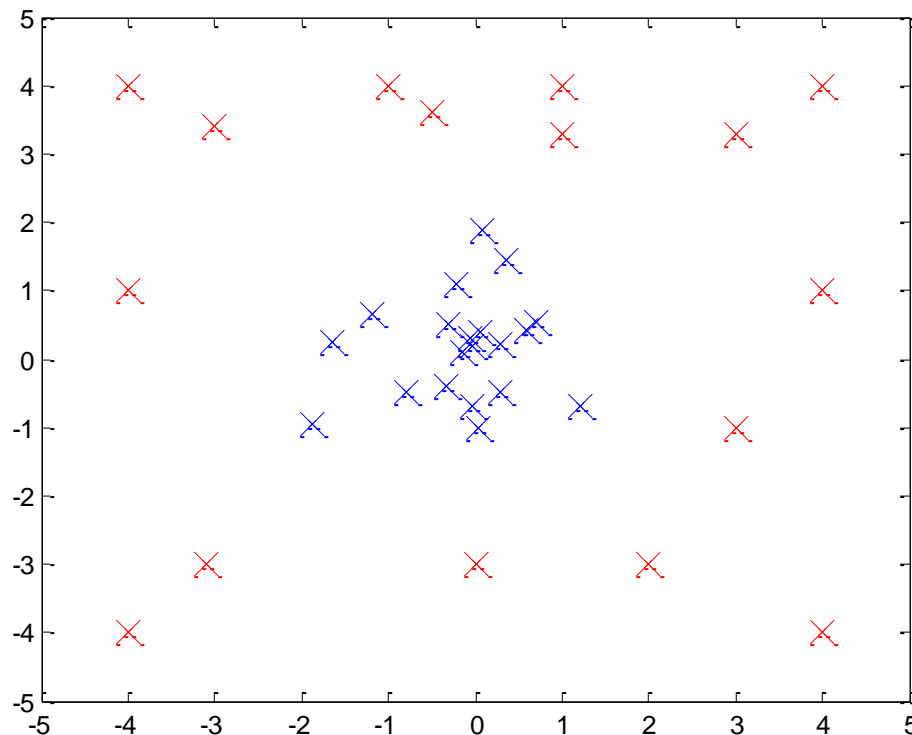
Regularização

Regressão Logística

Classificador Não Linear

- ▶ Como fazer essa classificação

- ▶ $x = [x_1 \ x_2]$

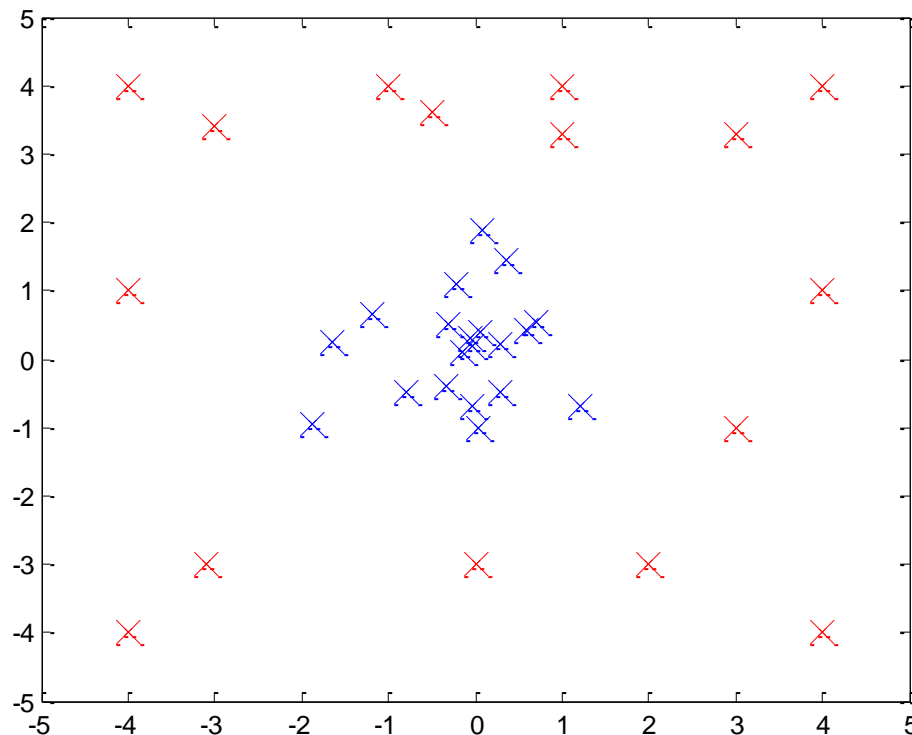


Classificador Não Linear

- ▶ Como fazer essa classificação

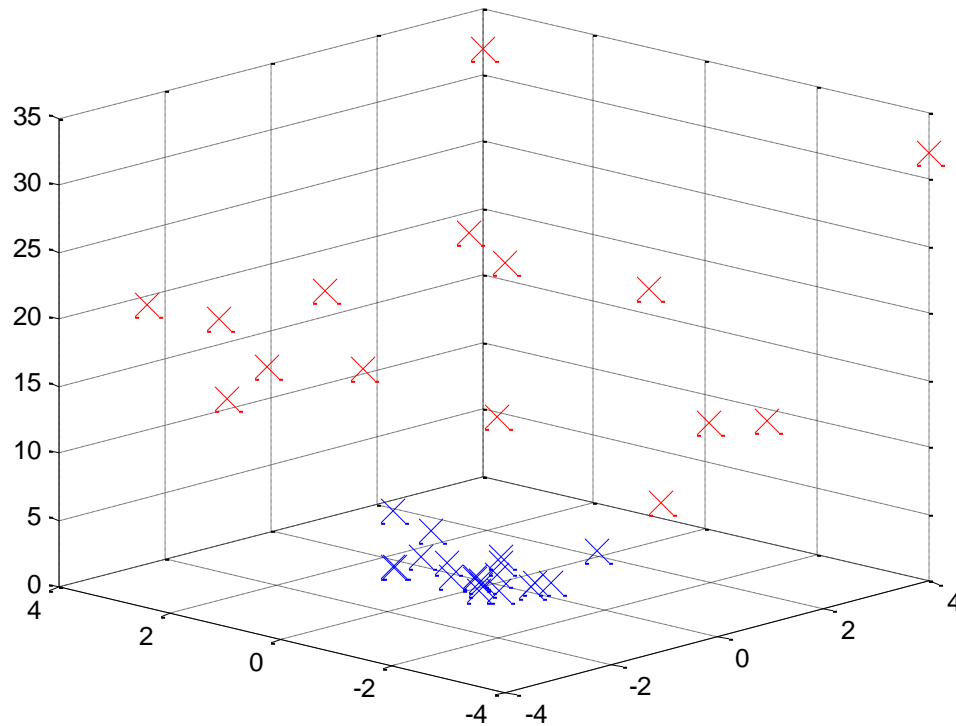
- ▶ $x = [x_1 \ x_2]$

- ▶ Criar um novo atributo - $x_3 = x_1^2 + x_2^2$



Classificador Não Linear

- Como fazer essa classificação



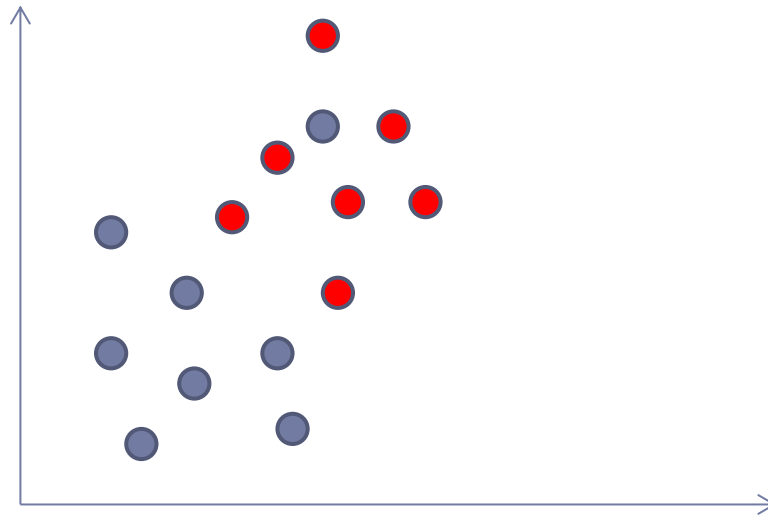
Overfitting

- ▶ Modelo se ajusta demasiadamente aos dados utilizados para encontrar os parâmetros



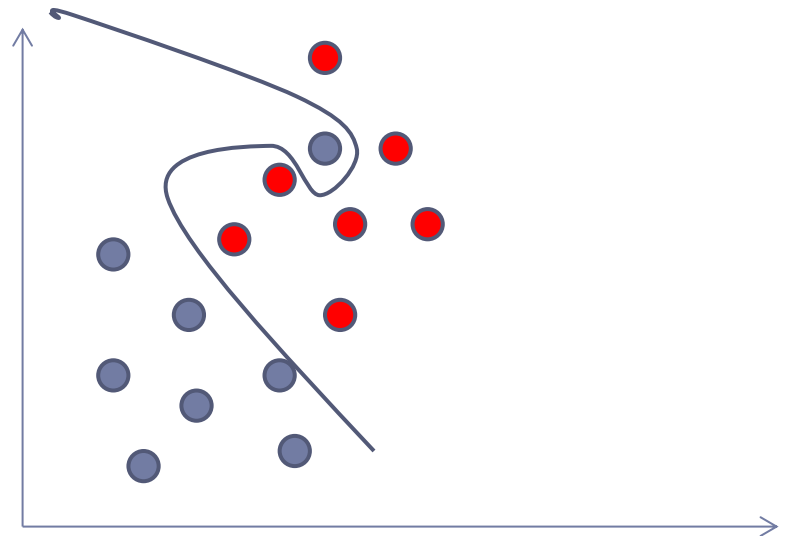
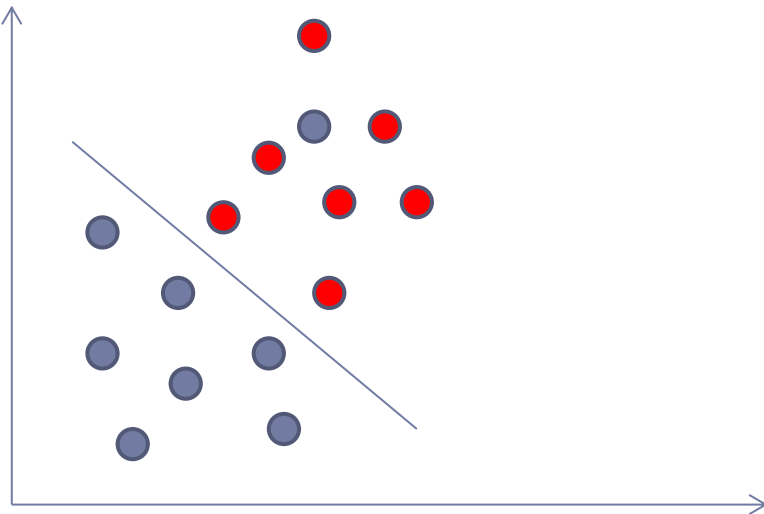
Overfitting

- ▶ Modelo se ajusta demasiadamente aos dados utilizados para encontrar os parâmetros



Overfitting

- ▶ Modelo se ajusta demasiadamente aos dados utilizados para encontrar os parâmetros



Tamanho do Modelo

- ▶ Como ajustar o tamanho do modelo ?

- ▶ Modelo

- ▶ $\bar{y}_i = \frac{1}{1+e^{-w^T x_i}}$

- ▶ \mathbf{x} pode ter dimensão alta



Tamanho do Modelo

- ▶ Como ajustar o tamanho do modelo ?

- ▶ Modelo

- ▶ $\bar{y}_i = \frac{1}{1+e^{-w^T x_i}}$

- ▶ \mathbf{x} pode ter dimensão alta

- ▶ Regularização

- ▶ Diminuir os pesos (w) associados às características (x) que influenciam menos no resultado



Regularização

- ▶ Função de custo

- ▶ $J(\mathbf{w}) = \frac{1}{2n} \sum_{i=1}^n -y_i \ln(\bar{y}_i) - (1 - y_i) \ln(1 - \bar{y}_i)$



Regularização

- ▶ Função de custo

- ▶ $J(\mathbf{w}) = \frac{1}{2n} \sum_{i=1}^n -y_i \ln(\bar{y}_i) - (1 - y_i) \ln(1 - \bar{y}_i)$

- ▶ Nova função de custo

- ▶ $J(\mathbf{w}) = \frac{1}{2n} [\sum_{i=1}^n -y_i \ln(\bar{y}_i) - (1 - y_i) \ln(1 - \bar{y}_i) + \lambda \sum_{j=1}^m w_j^2]$



Regularização

- ▶ Função de custo

- ▶ $J(\mathbf{w}) = \frac{1}{2n} \sum_{i=1}^n -y_i \ln(\bar{y}_i) - (1 - y_i) \ln(1 - \bar{y}_i)$

- ▶ Nova função de custo

- ▶ $J(\mathbf{w}) = \frac{1}{2n} [\sum_{i=1}^n -y_i \ln(\bar{y}_i) - (1 - y_i) \ln(1 - \bar{y}_i) + \lambda \sum_{j=1}^m w_j^2]$

- ▶ Regra de aprendizado

- ▶ Gradiente descendente

- $w = w - \alpha \frac{\partial J}{\partial w}$



Regularização

- ▶ Função de custo

- ▶ $J(\mathbf{w}) = \frac{1}{2n} [\sum_{i=1}^n -y_i \ln(\bar{y}_i) - (1 - y_i) \ln(1 - \bar{y}_i) + \lambda \sum_{j=1}^m w_j^2]$

- ▶ Gradiente descendente

- ▶ $w = w - \alpha \frac{\partial J}{\partial w}$



Regularização

- ▶ Função de custo

- ▶ $J(\mathbf{w}) = \frac{1}{2n} [\sum_{i=1}^n -y_i \ln(\bar{y}_i) - (1 - y_i) \ln(1 - \bar{y}_i) + \lambda \sum_{j=1}^m w_j^2]$

- ▶ Gradiente descendente

- ▶ $w = w - \alpha \frac{\partial J}{\partial w}$

- ▶ $\frac{\partial J}{\partial w} = -\frac{1}{n} \sum_{i=1}^n (\bar{y}_i - y_i) \mathbf{x}_i + \frac{\lambda}{n} w_j$



Regularização

- ▶ Função de custo

- ▶ $J(\mathbf{w}) = \frac{1}{2n} [\sum_{i=1}^n -y_i \ln(\bar{y}_i) - (1 - y_i) \ln(1 - \bar{y}_i) + \lambda \sum_{j=1}^m w_j^2]$

- ▶ Gradiente descendente

- ▶ $w = w - \alpha \frac{\partial J}{\partial w}$

- ▶ $\frac{\partial J}{\partial w} = -\frac{1}{n} \sum_{i=1}^n (\bar{y}_i - y_i) \mathbf{x}_i + \frac{\lambda}{n} w_j$

- ▶ Regra de Aprendizado

- ▶ $w_0 = w_0 + \alpha [\frac{1}{n} \sum_{i=1}^n e_i \mathbf{x}_{i0}]$

- ▶ $w_j = w_j + \alpha [\frac{1}{n} \sum_{i=1}^n e_i \mathbf{x}_{ij} - \frac{\lambda}{n} w_j]$



Regularização

- ▶ Gradiente descendente estocástico

- ▶ $w_0 = w_0 + \alpha e_i x_{i0}$

- ▶ $w_j = w_j + \alpha [e_i x_{ij} - \lambda w_j]$



Utilização do conjunto de dados
para treinamento e avaliação

Conjunto de Dados

- ▶ Conjunto de Treinamento e Conjunto de Testes (Holdout)



Conjunto de Dados

- ▶ Conjunto de Treinamento e Conjunto de Testes (Holdout)
 - ▶ Poucos dados



Conjunto de Dados

- ▶ Conjunto de Treinamento e Conjunto de Testes (Holdout)
 - ▶ Poucos dados
- ▶ Validação Cruzada
 - ▶ K-fold
 - ▶ Leave one out



K-fold

- ▶ Dividir o conjunto original em k subgrupos (*folds*) de mesmo tamanho
- ▶ Um *fold* é escolhido para teste
- ▶ Outros são utilizados para treinamento
- ▶ Procedimento é repetido k vezes
- ▶ Resultado é desempenho médio



Leave one out

- ▶ Caso particular do K-fold onde k é igual ao número de dados



Conjunto de Dados

- ▶ Conjunto de Treinamento e Conjunto de Testes (Holdout)
- ▶ Muitos dados



Conjunto de Dados

- ▶ Conjunto de Treinamento e Conjunto de Testes (Holdout)
- ▶ Muitos dados
 - ▶ Três conjuntos (treinamento, validação e teste)



Conjunto de Dados

- ▶ Conjunto de Treinamento e Conjunto de Testes (Holdout)
- ▶ Muitos dados
 - ▶ Três conjuntos (treinamento, validação e teste)
 - ▶ Ajuste dos parâmetros
 - ▶ Escolha dos hiper-parâmetros
 - ▶ Avaliação do desempenho



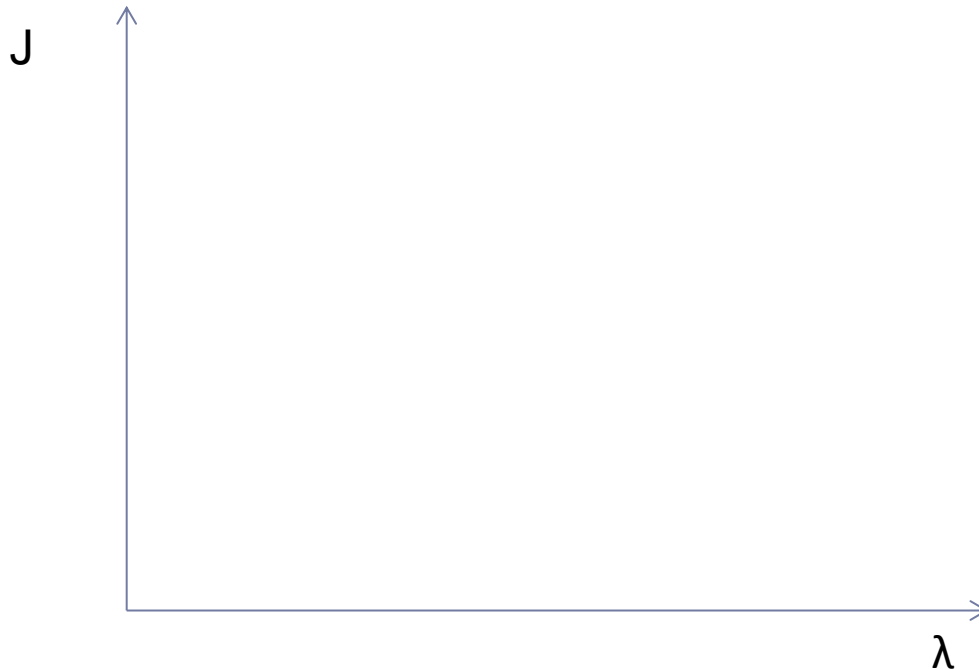
Conjunto de Dados

- ▶ Escolher um conjunto de valores para λ
- ▶ Treinar modelos com cada valor de λ
- ▶ Calcular o desempenho de cada modelo no conjunto de validação.
- ▶ Escolher o modelo com menor erro
- ▶ Obter o desempenho do modelo escolhido para o conjunto de teste



Conjunto de Dados

- ▶ Desempenho em cada conjunto



Problemas em um método de ML

- ▶ Mais dados
- ▶ Menos atributos
- ▶ Mais atributos
- ▶ Maior λ
- ▶ Menor λ





Dúvidas ?