Aprendizado de Máquina

Classificadores Lineares

Luiz Eduardo S. Oliveira

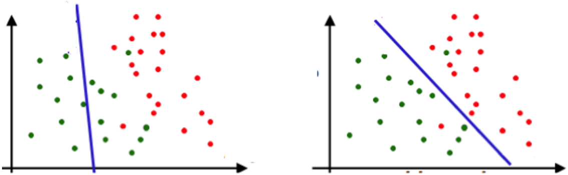
Universidade Federal do Paraná
Departamento de Informática
http://web.inf.ufpr.br/luizoliveira

Introdução

- A combinação linear das características é utilizada para tomar uma decisão
- Classificadores de rápido treinamento que em muitos casos atingem bom desempenho.

$$y = f(\vec{w} \cdot \vec{x} + b) = f(\sum_j w_j x_j + b)$$

- Durante o treinamento, os algoritmos desta categoria devem encontrar ($\vec{w}$) que melhor separa os dados.



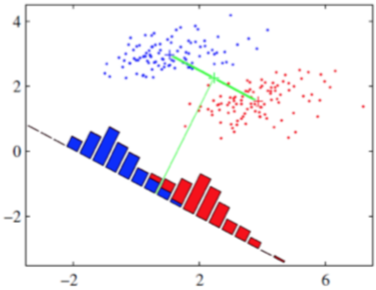
- As fronteiras de decisão são lineares (reta, plano, hiperplano)

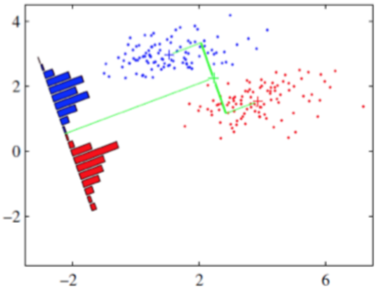
Introdução

- Existem duas classes de métodos para determinar os parâmetros de um classificador.
- **Generativos:** Levam em consideração da distribuição dos dados $P(\vec{x}|classe)$ (função de densidade condicional). Ou seja, leva em consideração como os dados foram gerados para então realizar a classificação
  - Linear Discriminant Analysis
- **Discriminativos:** A distribuição dos dados não é levada em consideração.
  - Regressão logistica
  - Perceptron
  - SVM

LDA

- LDA tenta encontrar uma transformação linear através da maximização da distância entre-classes e minimização da distância intra-classe.
- O método tenta encontrar a melhor direção de maneira que quando os dados são projetados em um plano, as classes possam ser separadas

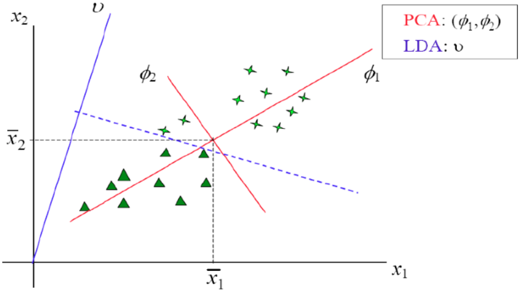




LDA

PCA vs LDA

- PCA encontra os principais vetores que descrevem os dados.
- LDA Encontrar o plano que separa os dados.
- Baseados no mesmo conceito de redução de dimensionalidade.



LDA

Exemplo: Considere a seguinte base de dados de controle de qualidade. x contém as duas medidas utilizadas na inspeção enquanto y indicam se o produto passou (+1) e não passou (-1) no teste, respectivamente.

x_1	x_2	y_i
2.95	6.63	+1
2.53	7.49	+1
3.57	5.65	+1
3.16	5.47	+1
2.58	4.46	-1
2.16	6.62	-1
3.27	3.52	-1

Construa um classificador LDA e classifique o $x_k = [2.81, 5.46]$

LDA

- 1. Encontrar os centroides de cada classe e o vetor médio.
  - $\mu_1 = [3.05, 6.38]$
  - $\mu_2 = [2.67, 4.73]$
  - $\mu = [2.88, 5.67]$
- 2. Normalizar os dados diminuindo o vetor μ de x
  - $x_1^0 = [[0.060, 0.951], [-0.357, 2.109], [0.679, -0.025], [0.269, -0.209]]$
  - $x_2^0 = [[-0.305, -1.218], [-0.732, 0.547], [0.386, -2.155]]$
- 3. Encontrar as matrizes de covariância para cada classe

$$c_1 = \begin{bmatrix} 0.166 & -0.192 \\ -0.192 & 1.149 \end{bmatrix} c_2 = \begin{bmatrix} 0.259 & -0.286 \\ -0.286 & 2.142 \end{bmatrix}$$

LDA

- 4. Encontrar a matriz conjunta de covariância para os pares (r,s)

$$C(r,s) = \sum_{i=1}^{k} P(i) \times c_i(r,s)$$

  - $C(0,0) = \frac{4}{7}0.166 + \frac{3}{7}0.259 = 0.206$
  - $C(1,1) = \frac{4}{7}1.349 + \frac{3}{7}2.142 = 1.689$

- Desta forma temos a matriz C abaixo

$$C = \begin{bmatrix} 0.206 & -0.233 \\ -0.233 & 1.689 \end{bmatrix}$$

- 5. Calcular a inversa de C

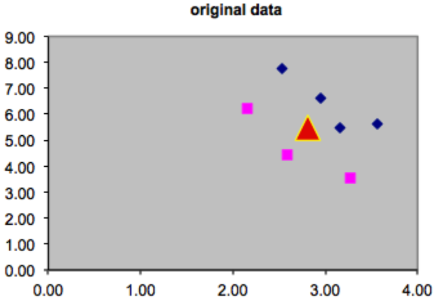
$$C^{-1} = \begin{bmatrix} 5.745 & -0.791 \\ 0.791 & 0.701 \end{bmatrix}$$

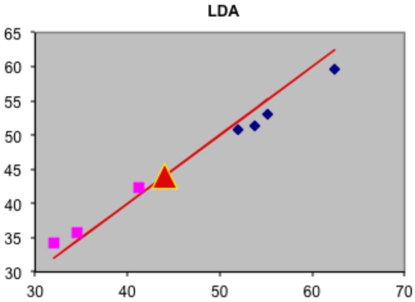
LDA

- 6. Finalmente a função discriminate é dada por

$$f_i = \mu_i C^{-1} x_k^T - \frac{1}{2}\mu_i C^{-1}\mu_i^T + \ln(P(i))$$

- Atribuimos o padrão x_k a classe i que maximiza f_i



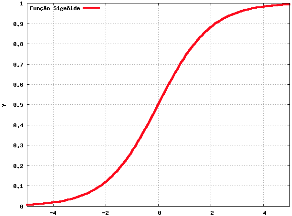


Regressão Logística

- Em alguns casos é interessante indicar ao usuário, além da classe, qual é a probabilidade de que um exemplo qualquer de teste pertença a uma determinada classe.
- Classificadores lineares atribuem somente a classe, ou seja, +1 ou -1.
- Uma alternativa consiste em adicionar uma função de ativação no nosso modelo linear

$$y = \sigma(w \cdot x + b)$$

- A função sigmoid σ é chamada de função "squashing" ou "logistic". Daí o nome, regressão logística, pois converte um valor real entre 0 e 1.



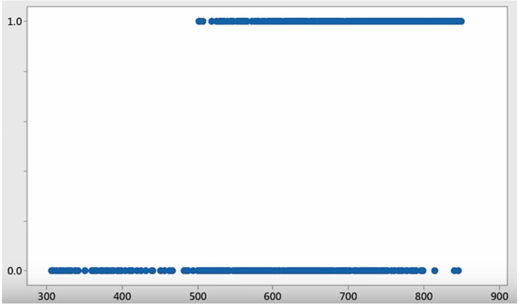
Regressão Logística

Exemplo

- Considere que alguém queira comprar uma casa e está na busca por um financiamento.
- Os bancos mantém para cada cliente um score, que varia de 300 a 850
- Suponha que um dado cliente tem um score de 720 e gostaria de saber qual é a probabilidade de ter seu credito aprovado pela instituição financeira.
- Esse cliente encontrou ainda dados de 1000 clientes com seus respectivos escores, que tiveram seus pedidos aprovados ou não.
- Notem que nesse caso a variável y é dicotômica
  - Aprovado (1)
  - Negado (0)

Regressão Logística

creditScore	approved
655	0
692	0
681	0
663	1
688	1
693	1
699	0
699	1
683	1
698	0
655	1
703	0
704	1
745	1
702	1



- Note que a distribuição dos dados é binomial.
- O modelos de regressão lineares não funcionam nesse caso.

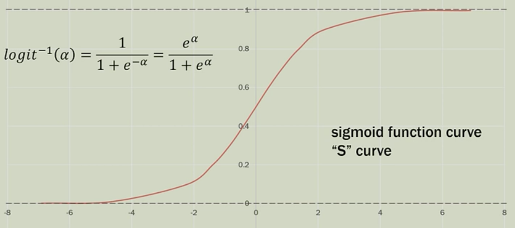
Regressão Logística

A regressão logística tem como objetivo:

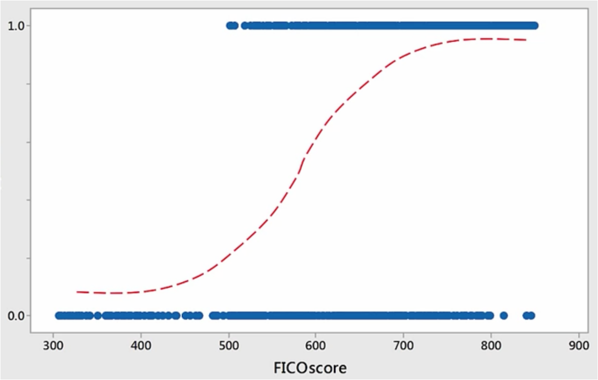
- Modelar a probabilidade de um evento ocorrer dependendo dos valores das variáveis independentes.
- Estimar a probabilidade de um evento ocorrer (e também de não ocorrer) para uma dada observação.
- Pode ser usada como um classificador, atribuindo uma classe ao padrão de entrada.
  - No nosso exemplo, credito aprovado ou reprovado.

Regressão Logística

- A variável dependente na regressão logística segue a distribuição de Bernoulli.
- Distribuição discreta de espaço amostral $\{0, 1\}$ que tem probabilidade de sucesso p e falha $q = 1 - p$
- Na regressão logística estimamos p para qualquer combinação linear das variáveis independentes.
- Isso pode ser alcançado ajustando o seguinte modelo (função logistica)
- MLE (Maximum Likehood Estimation) é usada para estimar os coeficiente do modelo.



Regressão Logística



- Esse modelo diz basicamente que a probabilidade de conseguir crédito sobre em função do score