



# Aprendizado Automático



João Paulo Pordeus Gomes

# Redução de Dimensionalidade

# Classificação e Regressão

---

- ▶ Diversos problemas reais têm muitos atributos
  - ▶ Reconhecimento de faces
  - ▶ Classificação de textos
  - ▶ Reconhecimento de impressões digitais
  - ▶ ...



# Classificação e Regressão

---

- ▶ Diversos problemas reais têm muitos atributos
  - ▶ Reconhecimento de faces
  - ▶ Classificação de textos
  - ▶ Reconhecimento de impressões digitais
  - ▶ ...
- ▶ Overfitting



# Classificação e Regressão

---

- ▶ Diversos problemas reais têm muitos atributos
  - ▶ Reconhecimento de faces
  - ▶ Classificação de textos
  - ▶ Reconhecimento de impressões digitais
  - ▶ ...
- ▶ Overfitting
- ▶ Custo computacional
- ▶ Custo de armazenamento



# Redução de Dimensionalidade

---

- ▶ Duas estratégias mais comuns
  - ▶ Seleção de atributos
  - ▶ Combinação de atributos



# Seleção de atributos

---

- ▶ Duas abordagens
  - ▶ Usar algum critério para definir o quão bom é um tributo (Filter)
  - ▶ Selecionar subgrupos de atributos e ver o resultado na classificação/regressão (Wrapper)



# Filter

---

- ▶ Fisher Score

- ▶ Toma cada um dos atributos e calcula

$$V(i) = \frac{(\mu_{(+)}(i) - \mu_{(-)}(i))^2}{\sigma_{(+)}^2(i) + \sigma_{(-)}^2(i)}$$

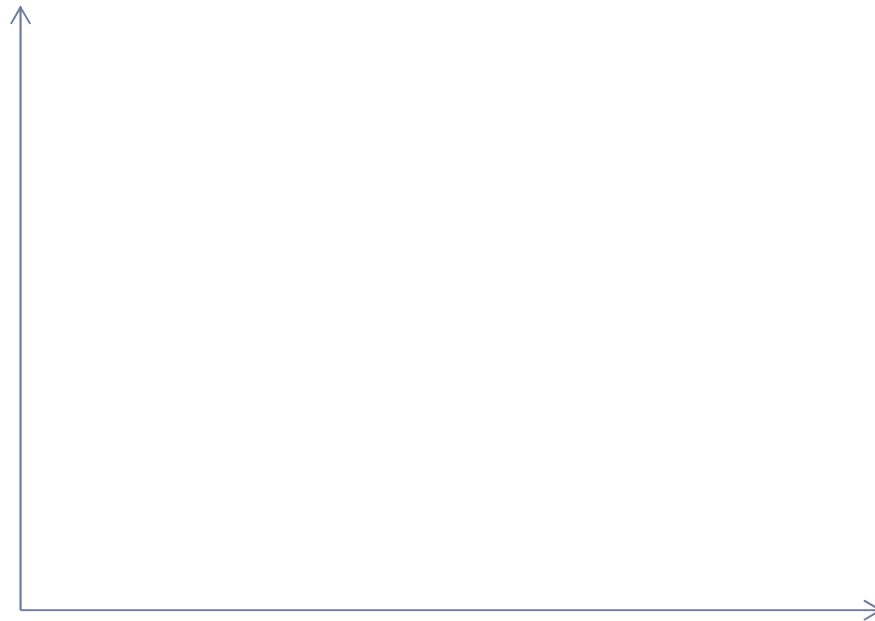




# Filter

---

- ▶ Fisher Score



# Wrapper

---

- ▶ Problema de otimização combinatória
- ▶ Escolher um subconjunto das variáveis originais que fornece o melhor resultado para classificação



# Wrapper

---

- ▶ Problema de otimização combinatória
- ▶ Escolher um subconjunto das variáveis originais que fornece o melhor resultado para classificação
- ▶ Pode utilizar uma estratégia gulosa (*greedy*)



# Combinação de atributos

---

- ▶ Combinar os atributos originais de forma a transformá-los em vetores de atributos de menos dimensão
- ▶ Como combinar ?
  - ▶ Linearmente
  - ▶ Não linearmente
- ▶ Qual critério usar para combinar ?



# Análise de Componentes Principais

---

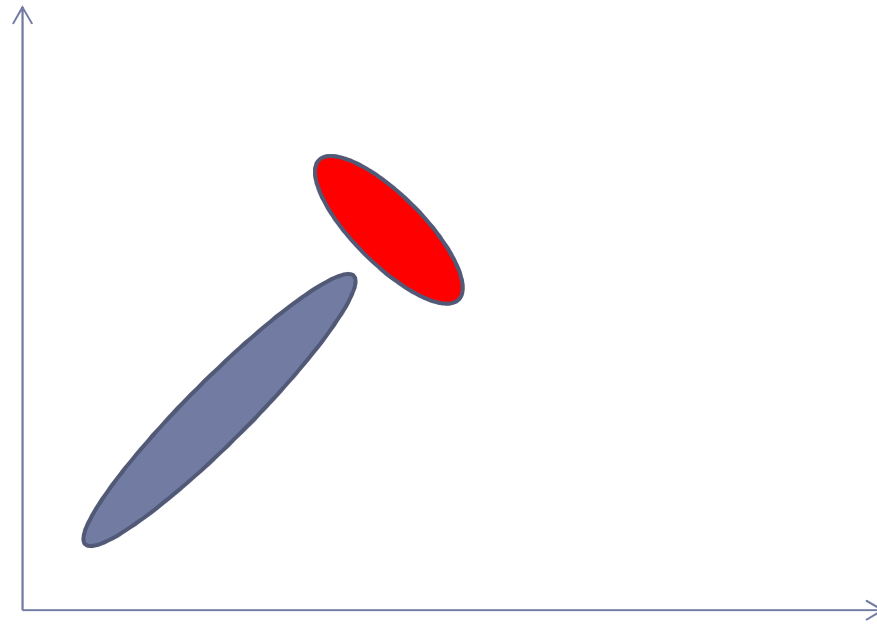
- ▶ Principal Component Analysis (PCA)
- ▶ Método de combinação linear dos atributos para a geração de novos atributos.
- ▶ Redução de dimensão se dá pela seleção de um subgrupo destes atributos



# PCA

---

## ► Idéia



# PCA

---

- ▶  $Ax = x'$
- ▶ Vamos definir  $u$  como uma parte de  $A$



# PCA

---

- ▶  $Ax = x'$
- ▶ Vamos definir  $u$  como uma parte de  $A$
- ▶ Temos que encontrar  $u$ , tal que:
  - ▶  $\sigma^2 = \frac{1}{N-1} \sum_{i=1}^N (x' - \mu) (x' - \mu)^T$
- ▶ Seja o maior possível





# PCA

---

- ▶  $Ax = x'$
- ▶ Vamos definir  $u$  como uma parte de  $A$
- ▶ Temos que encontrar  $u$ , tal que:
  - ▶  $\sigma^2 = \frac{1}{N-1} \sum_{i=1}^N (x' - \mu) (x' - \mu)^T$
- ▶ Seja o maior possível
- ▶ Sob certas restrições
- ▶  $u^t u = 1$



# PCA

---

$$\blacktriangleright \max \frac{1}{N-1} \sum_{i=1}^N (x' - \mu) (x' - \mu)^T$$

s.t

$$\blacktriangleright u^t u = 1$$



# PCA

---

$$\blacktriangleright \max \frac{1}{N-1} \sum_{i=1}^N (x') (x')^T$$

s.t

$$\blacktriangleright u^t u = 1$$



# PCA

---

- ▶  $\max \frac{1}{N-1} \sum_{i=1}^N (x') (x')^T$

s.t

- ▶  $u^t u = 1$

- ▶ Como:

- ▶  $\max \frac{1}{N-1} \sum_{i=1}^N (x') (x')^T$

- ▶  $= \max \frac{1}{N-1} \sum_{i=1}^N (ux) (ux)^T$



# PCA

---

- ▶  $\max \frac{1}{N-1} \sum_{i=1}^N (x') (x')^T$

s.t

- ▶  $u^t u = 1$

- ▶ Como:

- ▶  $\max \frac{1}{N-1} \sum_{i=1}^N (x') (x')^T$

- ▶  $= \max \frac{1}{N-1} \sum_{i=1}^N (ux) (ux)^T$

- ▶  $= \max u^T \sum u$



# PCA

---

►  $= \max u^T \Sigma u$

s.t

►  $u^t u = 1$



# PCA

---

▶  $= \max u^T \Sigma u$

s.t

▶  $u^t u = 1$

▶ Multiplicadores de Lagrange



# PCA

---

▶  $= \max u^T \Sigma u$

s.t

▶  $u^t u = 1$

▶ Multiplicadores de Lagrange

▶  $L = u^T \Sigma u - \lambda(u^t u - 1)$





# PCA

---

▶  $= \max u^T \Sigma u$

s.t

▶  $u^t u = 1$

▶ Multiplicadores de Lagrange

▶  $L = u^T \Sigma u - \lambda(u^t u - 1)$

▶  $\frac{\partial L}{\partial u} = 0$

▶  $\frac{\partial L}{\partial \lambda} = 0$



# PCA

---

▶  $= \max u^T \Sigma u$

s.t

▶  $u^t u = 1$

▶ Multiplicadores de Lagrange

▶  $L = u^T \Sigma u - \lambda(u^t u - 1)$

▶  $\Sigma u = \lambda u$

▶  $u^t u = 1$



# PCA

---

- ▶ Como resolver

- ▶  $\Sigma u = \lambda u$



# PCA

---

- ▶ Como resolver
  - ▶  $\Sigma u = \lambda u$
- ▶ Problema de Autovetores e autovalores



# PCA

---

- ▶ Como resolver
  - ▶  $\Sigma u = \lambda u$
- ▶ Problema de Autovetores e autovalores
- ▶ Para achar  $u$ , basta calcular os autovetores da matriz  $\Sigma$
- ▶  $\lambda$  são os autovalores associados
- ▶ Os autovalores correspondem a variância explicada em cada novo atributo



## PCA na prática

---

- ▶ Toma os dados de entrada
- ▶ Subtrai as médias
- ▶ Calcula a matriz de covariância
- ▶ Calcula os autovalores e autovetores
- ▶ Escolhe os  $k$  maiores autovalores
- ▶ Utiliza os  $k$  autovetores correspondentes para criar  $k$  novos atributos



Dúvidas ?