

Exercício 1 - MC886

João Pedro Fabiano Franco - RA 146641

Código

```
import pandas
import numpy
from sklearn.decomposition import PCA
from sklearn.linear_model import LogisticRegression
from sklearn.lda import LDA

# 1. Faça o PCA dos dados (sem a última coluna). Se voce quiser que os dados transformados tenham 80\% da variância original, quantas dimensões do PCA vc precisa manter?
data = pandas.read_csv("data1.csv") # le o arquivo .csv
dataWithoutClass = data.drop(labels='clase', axis=1) # tira a coluna de classes para PCA

pca = PCA(n_components=0.8) # PCA com dados que tenham 80% da variancia original
dataTransformed = pca.fit_transform(dataWithoutClass) # aplica PCA

print(pca.n_components_) # resultado

# 2. Treine uma regressão logística no conjunto de treino dos dados originais e nos dados transformados. Qual a taxa de acerto no conjunto de teste nas 2 condições (sem e com PCA)?
dataTraining = data[:200] # conjunto de treino dos dados originais
dataTest = data[200:] # conjunto de teste dos dados originais
dataTransformedTraining = dataTransformed[:200] # conjunto de treino dos dados transformados
dataTransformedTest = dataTransformed[200:] # conjunto de teste dos dados originais
dataClassTraining = dataTraining.ix[:, 166] # classes do conjunto de treino
dataClassTest = dataTest.ix[:, 166] # classes do conjunto de teste

logReg1 = LogisticRegression()
logReg1.fit(dataTraining, dataClassTraining) # aplica regressao logistica no conjunto de treino dos dados originais
print("Dados originais:", logReg1.score(dataTest, dataClassTest)) # taxa de acerto sobre o conjunto de teste dos dados originais

logReg2 = LogisticRegression()
logReg2.fit(dataTransformedTraining, dataClassTraining) # aplica regressao logistica no conjunto de treino dos dados transformados
print("Dados transformados:", logReg2.score(dataTransformedTest, dataClassTest)) # taxa de acerto sobre o conjunto de teste dos dados transformados

# 3. Treine o LDA nos conjuntos de treino com e sem PCA e teste nos respectivos conjuntos de testes. Qual a acurácia nas 2 condições?
lda1 = LDA()
lda1.fit(dataTraining, dataClassTraining) # aplica LDA no conjunto de treino dos dados originais
print("Dados originais:", lda1.score(dataTest, dataClassTest)) # taxa de acerto sobre o conjunto de teste dos dados originais

lda2 = LDA()
lda2.fit(dataTransformedTraining, dataClassTraining) # aplica LDA no conjunto de treino dos dados transformados
print("Dados transformados:", lda2.score(dataTransformedTest, dataClassTest)) # taxa de acerto sobre o conjunto de teste dos dados transformados
```

Respostas

13

Código

1. Faça o PCA dos dados (sem a última coluna). Se voce quiser que os dados transformados tenham 80\% da variância original, quantas dimensões do PCA vc precisa manter?

2. Treine uma regressão logística no conjunto de treino dos dados originais e nos dados transformados. Qual a taxa de acerto no conjunto de teste nas 2 condições (sem e com PCA)?

3. Treine o LDA nos conjuntos de treino com e sem PCA e teste nos respectivos conjuntos de testes. Qual a acurácia nas 2 condições?

4. Qual a melhor combinação de classificador e PCA ou não?

Analizando os resultados, concluímos que o melhor classificador é com regressão logística. Com PCA, tanto regressão quanto LDA tem taxa de acerto parecidos, mas sem PCA, regressão se saiu melhor.

Assunto

Estilo Layout Mais

Fonte