Factorial Discriminant Analysis

1. General principle

2. Factorial Discriminant Analysis with Python

1 General principle

2 Factorial Discriminant Analysis with Python

Scientific question

- Data : n observations characterized by m quantitative variables (data matrix X) and one single qualitative variable (label Y)
- Goals :
  - describe : how can we discriminate the classes
  - classify : using explanatory variables, decide how classify a new observation

Factorial Discriminant Analysis vs PCA

- PCA maximizes the variance of projections on a convenient subspace
- Factorial Discriminant Analysis maximizes the difference between classes on a convenient subspace

1 General principle

2 Factorial Discriminant Analysis with Python

A basic example

```
import numpy as np
from sklearn.discriminant_analysis import
LinearDiscriminantAnalysis
X = np.array([[-1, -1], [-2, -1], [-3, -2], [1, 1],
[2, 1], [3, 2]])
y = np.array([1, 1, 1, 2, 2, 2])
clf = LinearDiscriminantAnalysis()
clf.fit(X, y)
```

Real data

The dataset used here is bank note authentication dataset publicly available in UCI machine learning repository.

https://archive.ics.uci.edu/ml/datasets/banknote+authentication#

```
import pandas as pd
columns =
["var","skewness","kurtosis","entropy","class"]
df = pd.read_csv("http://archive.ics.uci.edu/ml/
machine-learning-databases/00267/
data_banknote_authentication.txt",index_col=False,
names = columns)
df
```

Real data

We visualize the data

```
import matplotlib.pyplot as plt
f, ax = plt.subplots(1, 4, figsize=(10,3))
vis1 = sns.distplot(df["var"],bins=10, ax= ax[0])
vis2 = sns.distplot(df["skewness"],bins=10, ax=ax[1])
vis3 = sns.distplot(df["kurtosis"],bins=10, ax=
ax[2])
vis4 = sns.distplot(df["entropy"],bins=10, ax=ax[3])
```

Real data

We separate features and labels

```
df_values=df.values
df_values
df_data=df_values[:,:4]
df_label=df_values[:,4]
```

Real data

We define and fit the model

```
import sklearn
from sklearn.discriminant_analysis import
LinearDiscriminantAnalysis
lda = LinearDiscriminantAnalysis()
lda.fit(df_data,df_target)
```

Real data

We visualize

```
df_data_trans = lda.transform(df_data)
plt.plot(df_data_trans, df_target, 'r+')
plt.show()
```

Factorial Discriminant Analysis, how use it to classify?

- Goal : classify an new observation using explanatory variables
- Approach : assign the observation to the nearest class