

td7_sujet

November 15, 2021

Université de Bordeaux, Master Mention Informatique, 2021-2022

1 Analyse, classification et indexation des données: feuille 7

1.0.1 Réduction de dimension - Analyse linéaire discriminante (LDA)

Avant de commencer Dans cet exercice, on s'intéresse à l'application de l'analyse linéaire discriminante (LDA) à un corpus de données. Le but est de réduire la dimension tout en conservant une séparation entre les classes.

Questions préliminaires :

1. Quelle est la différence entre une ACP et une LDA ?
2. Quelle est la fonction objective que l'on cherche à minimiser quand on fait une LDA ?

1.1 Exercice 1.

Dans cet exercice, on considère le corpus sur la classification des vins. Ce corpus fait partie de l'ensemble des datasets du module `sklearn.datasets`. Il peut être chargé en invoquant la méthode `load_wine()`.

1.1.1 Préparation des données

1. Chargez le corpus, explorez-le.

[]:

2. Créez un DataFrame `data` contenant les variables indépendantes et mettez les classes dans une variable dépendante `target`. Quelle est la taille du corpus ? Combien de variables comporte-t-il ?

[]:

[]:

3. En utilisant la fonction `cocat`, constituez un seul corpus dataset contenant `data` et `target`. Faites un mélange des données

[]:

[]:

LDA “à la main” Dans un premier temps, nous allons appliquer la méthode d’analyse linéaire discriminante juste sur deux classes. Nous allons donc d’abord ne garder que les vins des classes 1 et 2. Nous allons également ne conserver que deux descripteurs alcohol et color_intensity.

1. Ecrivez les instructions permettant de faire ce filtrage

[]:

2. Quelles sont les étapes (théoriques) pour réaliser une LDA ?
3. Application numérique : appliquer les étapes de la questions précédente pour réaliser une LDA. Quel est le vecteur directeur du meilleur axe de projection selon la méthode LDA ?

[]:

3. Visualiser le résultat. Donnez les deux graphiques, avant et après la projection.

[]:

Avant la projection :

[]:

Après la projection :

[]:

1.1.2 LDA avec sklearn

Réalisez une LDA en utilisant la bibliothèque sklearn.discriminant_analysis. Comparez le résultat avec la projection que vous avez obtenu dans la section précédente.

[]:

[]:

[]:

[]:

1.1.3 LDA sur tout le corpus

Nous allons à présent appliquer la LDA sur tout le corpus. L'objectif est d'observer l'impact de la projection à la fois sur l'efficacité (accuracy) de la classification et sur le temps d'apprentissage.

1. Rechargez les données et faites votre LDA.

[]:

[]:

2. Quel est le ration de la variance expliqué par les axes obtenus ?

[]:

3. Visualiser le résultat.

[]:

1.1.4 Classification avant/après réduction

1. En utilisant un classifieur bayésien MAP, comparez les résultats obtenus avec les échantillons bruts et les résultats obtenus avec les échantillons projetés sur les axes fournis par la LDA. Comparez également les temps d'entraînement.

[]:

[]:

[]:

[]:

2. Même question avec un k -nn.

[]:

[]:

[]:

[]:

1.2 Exercice 2. (A vous de jouer !)

Dans cet exercice, nous allons travailler avec le même corpus que le précédent TD : les données sur le cancer du sein. Le corpus peut être chargé par l'instruction `load_breast_cancer` de la bibliothèque

`sklearn.datasets`.

Comparer les effets de l'ACP et de la LDA sur la classification de ce corpus.