パターン情報処理　課題

2019/12/17

目加田慶人

以下の課題から必須課題を含めて少なくとも２題を選択し，その処理を実現するプログラムを作成せよ．レポートには，その処理の主要部分，出力結果とそれらに対する説明と考察を加えよ．

締切り：1 月 21 日

提出物：
レポート本体は，MaNaBo で以下の 2 点を提出すること．

１．レポート本体（ワードまたは pdf）
２．ソースファイル(kadai1.cpp, kadai2.m, kadai3.py, kadai4.ipynb など)や自分で生成したデータ（csv 形式）を zip 等でまとめたファイル

注意事項：

プログラミング言語は, 基本的には Python とする. 各自でデータセットをコンバートすれば，Matlab や C，C++を利用してもよい.

１．【必須問題】競争問題（この問題のみ言語は python に限定します）

以下の条件でデータを生成し，これらを授業で教えた分類器で分類し，テストデータに対する正解率（accuracy_score）および，テストデータと分類境界線がわかる図を示せ．またテストデータに対する混同行列を示せ.

```
from sklearn.datasets import make_blobs
from sklearn.model_selection import train_test_split
X, y = make_blobs(random_state=122, n_samples=450, n_features=2,　cluster_std=1.8, centers=4)
X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=1)
```

使ってよい分類器は，k 最近傍識別，ニューラルネットワーク，ランダムフォレスト，線形 SVM，カーネル SVM，判別分析とする．正解率の高い人を高く評価します.

２．最近傍識別器

0 から 9 の 10 種類の文字の手書き数字を縦横それぞれ 8 画素にした画像データである，Optical recognition of handwritten digits dataset に対して, 適当にデータを分割して最近傍識別機で分類せよ．正しく分類された結果の例，および分類誤りの例をしめせ．混同行列を示し，そこから正答率を計算せよ．これらの実験結果を考察せよ.

３．パラメータ推定とベイズ決定

sample1.csv の各クラスが正規分布に従うとする．各クラスの事前確率，平均ベクトルと共分散行列を求めこれを示せ．またここで求めた値を用い，ベイズ決定則(p.50)により sample2.csv を分類し，その結果を考察せよ．

４．特徴ベクトルの正規化

特徴ベクトルの正規化について，その効果を 400 字以内でまとめよ．（中心化や標準化と呼ばれる処理を含めても良いです）

５．次元削減

scikit learn に用意されている iris データに対して KL 展開（主成分分析）による次元削減により 4 次元を 2 次元にせよ．削減された 2 次元空間で各クラスのデータの分布をプロットしたものと，特定の 2 軸を選択してプロットしたものとの比較を考察せよ．

６．次元削減，線形識別

scikit learn で two moons data を適当に生成し，これに対して Fisher の線形識別をおこなえ．結果は，新しく得られた 1 次元軸に対する 2 クラスのデータのヒストグラムを示すこと．

■データフォーマットについて

こちらから提供するデータのフォーマットは以下の通りである．自身の環境に合わせて適当に形式を変換して利用すること．

・テキスト形式
・1 行目に特徴ベクトルの次元数
・2 行目にクラス数
・3 行目以降は，特徴ベクトルの各次元が順にカンマ(,)で区切られ，最後にクラス番号

つまり，特徴ベクトルの次元数を d，クラス数を c，総データ数を n とすると，このファイルは全体で n+2 行で，3 行目から n+2 行目までは d+1 個の数字がカンマで区切られていることになる．