パターン情報学　プログラミングレポート課題

03-140299 東京大学機械情報工学科 3 年 和田健太郎

2015 年 1 月 22 日

1　課題 1

課題 1

2 クラス（$\omega 1$，$\omega 2$）の識別問題を考える．データは 2 次元とする．配布するデータセットの説明を以下に示す．

- Train1.txt，Train2.txt：$\omega 1$，$\omega 2$ に属する訓練データ集合．各データ数 50.
- Test1.txt，Test2.txt： $\omega 1$，$\omega 2$ に属するテストデータ集合．各データ数 20.

2 クラスで，2 次元のデータに対するウィドロー・ホフのアルゴリズムを実装し，訓練データから分離超平面を学習せよ．また，テストデータの識別率（全テストデータ数に対する正しく識別されたテストデータ数の比率）を求めよ．さらに，訓練データ，テストデータ，学習された識別面を図示せよ．

ウィドロー・ホフのアルゴリズムを初期の重みはランダムとし, 指定した回数だけ繰り返し重みの更新を行うように実装した．

2 次元の訓練データ 100 件を用いて識別器の学習を行い, 40 件のテストデータで性能を測定したところ, 0.875 という結果が出た．

また, 訓練データ, テストデータのそれぞれ 2 クラスと識別面を図示したものが図 1 である．

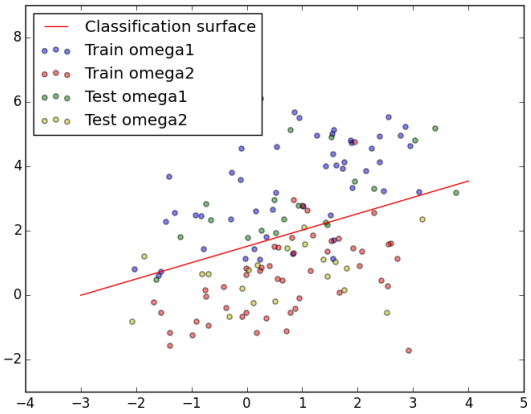

図 1: データおよび識別面

2　課題 2

課題 2

擬似逆行列を計算するプログラムを書き，課題 1 と同じ訓練データから分離超平面を学習せよ．また，テストデータの識別率を求めよ．クラスラベルについて，$\omega 1$ に属するものを 1，$\omega 2$ に属するものを-1 などとせよ．さらに，学習された識別面を課題 1 と同じ図に示せ．

擬似逆行列を数値計算ライブラリである numpy を利用して実装した．

$$A^{+} = (A^{T} \cdot A)^{-1} \cdot A^{T}$$

擬似逆行列を用いて訓練データに関して重みを計算し, テストデータによって識別性能を測定したところ, 1 と同様に 0.875 という結果だった．

訓練データ, テストデータおよび識別面を図示したものが図 2 で, 識別面の位置をウィドロー・ホフのアルゴリズムによるものと比べてみると, ほぼ同じ位置にあることがわかる．

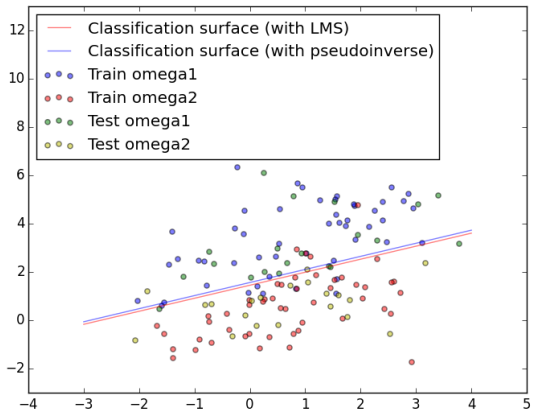

図 2: データおよび識別面

3 課題 3

課題 3

本課題も課題 1 と同じデータセットを利用する．

1. テストデータの集合を k 近傍法（kNN）を用いて識別することを考える． 訓練データに対して一つ抜き法（LOO: leave-one-out）により k の値を 1 から 10 まで変化させ，最適な k の値を求めよ．また，横軸に k， 縦軸に識別率としてグラフを作成せよ．

2. LOO により得られた k の値を用いてテストデータを識別せよ．そして，識別率を求めよ．

4 課題 4

課題 4

表にあるデータを利用する．また潜在的な確率密度分布は正規分布であるとする．P(ωi)=1/3 とする．表にあげた各クラスのデータセットは omega1.txt, omega2.txt，omega3.txt である．このとき次の問いに答えよ．

1. テスト点：(1, 2, 1)T , (5, 3, 2)T , (0, 0, 0)T , (1, 0, 0)T と各クラスの平均との間のマハラノビス距離を求めよ．

2. これらの点を識別せよ．

3. 次に P(ω1)=0.8 かつ P(ω2) = P(ω3)=0.1 と仮定し，テスト点をもう一度識別せよ