

レポート課題1

課題

各自作成した認識プログラムを実行して以下の問題に答えよ。データは与えられた2次元データ(2D-class1_A.dat、2D-class2_A.dat、2D-test_A.dat)を使用すること。なおレポートには必ず自分の作成したプログラムを添付すること。[発展課題]はレポート提出に際して必須ではないが、得点向上のため余裕があれば挑戦してもらいたい。

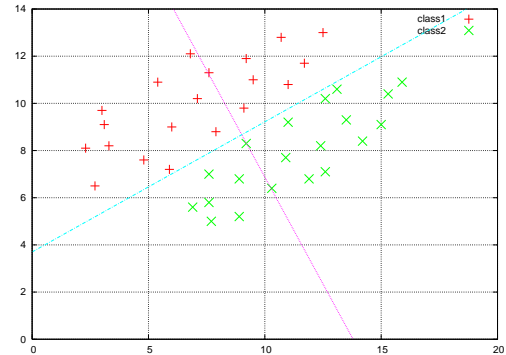
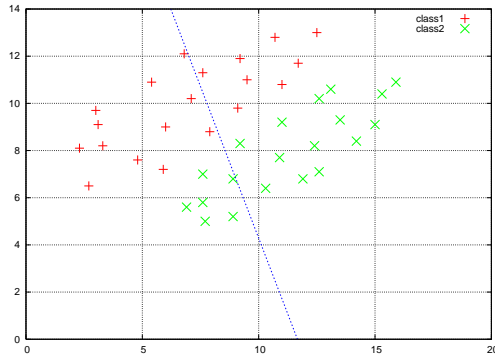
$$\text{※認識率は次の式で求める} \quad : \quad \text{認識率} [\%] = \frac{\text{認識正解データ数}}{\text{未知データ数}} \times 100$$

1. 判別評価基準 (ユークリッド・類似度・重みつきユークリッド距離)

- (a) Nearest Neighbor、ユークリッド距離、類似度、重みつきユークリッド距離それぞれによる認識プログラムをサンプルプログラム (sample1.c) を参考に作成・実行し、その結果を認識正誤表にまとめ、認識率を求めよ。
(※重みつきユークリッド距離においては使用した重み w_i を必ず明記すること)
- (b) 学習データ (2D-class1_A.dat、2D-class2_A.dat)、テストデータ (2D-test_A.dat(正解クラスごとに点のタイプを変えること)) の2次元散布図を作成せよ。その際に、学習データの平均もプロットし、学習データの散布図から、それぞれの評価基準における識別境界 (直線または区分的直線) を求め図示せよ。但し、Nearest Neighbor については、おおまかなもので良い。その結果より、各評価基準においてテストデータの誤認識が起こる領域について述べよ。
- (c) (a),(b) を、以下の学習・テストデータについても同様に行い、(a),(b) と比較せよ。
 - * set B : 2D-class1_B.dat、2D-class2_B.dat、2D-test_B.dat
 - * set C : 2D-class1_C.dat、2D-class2_C.dat、2D-test_C.dat
 - * set D : 2D-class1_D.dat、2D-class2_D.dat、2D-test_D.dat
- (d) (a),(b),(c) の結果から各判別評価基準についてまとめ、考察せよ。特に、Nearest Neighbor、ユークリッド距離、類似度において誤認識が発生するケースについて比較し、詳しく述べよ。
- (e) [発展課題] 重みつきユークリッド距離において、重みを様々に変化させ、そのときの認識率の変化をグラフに示せ。どのような重みのつけ方が有効であるか考察せよ。

2. 判別分析 (LDA) ・主成分分析 (PCA)

- (a) 判別分析 (LDA) ・主成分分析 (PCA) それぞれの手法を用いた認識プログラムを作成・実行し、その結果を認識正誤表にまとめ、認識率を求めよ。set A のみでよい。
(※判別評価基準はユークリッド距離を用いること)
- (b) (a) におけるプログラムによって得られた判別軸・主成分軸 (データを射影する軸) を学習データと共に下の図にならってプロットせよ。(※固有ベクトルの値より、直線 $y = ax + b$ の傾き a を求める。切片 b の値は図が見やすいように各自調整せよ。)



- (c) 1.(b) と同様に、それぞれの分析手法における識別境界を図示せよ。
- (d) (b)(c) の結果から各分析法について考察せよ。
- (e) [発展課題] 以上の実験において、良好な結果が得られなかったものを取り上げ、その原因を述べよ。また、良好な結果が得られると思われるケースについて、各自の考察を記せ。