

レポート課題のサンプル解答

正しくプログラムが組めていれば、以下のような値が出るはずである。

1. (a)

・ Nearest Neighbor

NUM	MET1	MET2	RES	ANS	o/x
1	2.26	11.45	1	1	o

・ ユークリッド距離

NUM	MET1	MET2	RES	ANS	o/x
1	8.40	36.09	1	1	o

・ 類似度

NUM	MET1	MET2	RES	ANS	o/x
1	1.00	0.94	1	1	o

・ 重み付きユークリッド距離

NUM	MET1	MET2	RES	ANS	o/x
1	77.98	83.07	1	1	o

※ただし各次元の重みを $w_1=1$, $w_2=90$ とした場合

1. (c)

SET B

・ Nearest Neighbor

NUM	MET1	MET2	RES	ANS	o/x
1	2.26	9.58	1	1	o

・ ユークリッド距離

NUM	MET1	MET2	RES	ANS	o/x
1	8.40	21.41	1	1	o

・ 類似度

NUM	MET1	MET2	RES	ANS	o/x
1	1.00	0.96	1	1	o

・ 重み付きユークリッド距離

NUM	MET1	MET2	RES	ANS	o/x
1	14.45	31.09	1	1	o

※ただし各次元の重みを $w_1=2$, $w_2=3$ とした場合

SET C

・ Nearest Neighbor

NUM	MET1	MET2	RES	ANS	o/x
1	2.26	11.45	1	1	o

・ ユークリッド距離

NUM	MET1	MET2	RES	ANS	o/x
1	8.40	36.09	1	1	o

・類似度

NUM	MET1	MET2	RES	ANS	o/x
1	1.00	0.88	1	1	o

・重み付きユークリッド距離

NUM	MET1	MET2	RES	ANS	o/x
1	24.66	26.27	1	1	o

※ただし各次元の重みを $w_1=0.1$, $w_2=9$ とした場合

SET D

・Nearest Neighbor

NUM	MET1	MET2	RES	ANS	o/x
1	2.26	11.45	1	1	o

・ユークリッド距離

NUM	MET1	MET2	RES	ANS	o/x
1	8.72	36.09	1	1	o

・類似度

NUM	MET1	MET2	RES	ANS	o/x
1	1.00	0.94	1	1	o

2. (a) 1つ目の未知データの識別結果

・判別分析

NUM	MET1	MET2	RES	ANS	o/x
1	0.49	2.55	1	1	o

・主成分分析

NUM	MET1	MET2	RES	ANS	o/x
1	3.61	36.09	1	1	o

判別分析・主成分分析についての捕捉

判別分析・主成分分析において固有値問題を解いた場合、有効な (固有値が 0 でない) 固有ベクトルは、以下の本数ずつ求まる。

- 判別分析 : $\min(\text{クラス数} - 1, \text{クラスごとの学習データ数} - 1)$ 本
- 主成分分析 : $\min(\text{元の次元数}, \text{全学習データ数} - 1)$ 本

2 クラス 2 次元データ - タの識別で学習データ数が 20 の場合は、判別分析で 1 本、主成分分析で 2 本の固有ベクトルが得られることになる。

判別分析では得られた固有ベクトル全てに射影することで識別を行うが、主成分分析では得られた固有ベクトルのうち、固有値の大きな順に任意の本数を選択し射影する。これにより元の次元数に比べて扱う次元数が小さくなるため、情報量が圧縮されることになる (次元圧縮)。

また、主成分分析において射影する固有ベクトル ϕ_1, ϕ_2, \dots , を元の次元数 d だけ選択した場合は、ユークリッド距離に等しくなる。

$$\mathbf{x} - \mathbf{m} = \sum_{i=1}^d \phi_i^T (\mathbf{x} - \mathbf{m})$$

但し、 \mathbf{x} : 未知データ、 \mathbf{m} : 平均ベクトル

レポート課題の場合は 2 次元データを扱っているため、得られた 2 本固有ベクトルのうち、固有値の大きな方に射影する。(2 本両方に射影した場合、ユークリッド距離の結果と等しくなるか確認してみると良い。)