

# パターン認識論レポート 2

b5tb2080 江 宇揚 (Jiang Yuyang)

Jul/15/2018

本レポートのプログラムと結果を全て github にアップロードしており、必要であればご確認ください。url: <https://github.com/huyuu/patternRecogReport2>

## 1 判別評価基準 (ユークリッド・類似度・重み付きユークリッド) :

### 1.1 (a) ユークリッド距離法と類似度を用いた結果

ユークリッド距離法と類似度を用いて文字画像認識プログラムを作成し、実行した結果、認識正誤表の一部および認識率を合わせて表 1 に示した。

表 1 ユークリッド距離法と類似度による正誤表及び認識率

| データ番号 | ユークリッド | 類似度   |
|-------|--------|-------|
| 0     | True   | True  |
| 1     | True   | True  |
| 2     | True   | True  |
| 3     | True   | True  |
| 4     | True   | True  |
| 5     | True   | True  |
| 6     | True   | True  |
| 7     | True   | True  |
| 8     | True   | True  |
| ..    | ...    | ...   |
| 73    | True   | True  |
| 74    | True   | True  |
| 75    | True   | True  |
| 76    | False  | False |
| 77    | True   | True  |
| 78    | True   | True  |
| 79    | True   | True  |
| 認識率 % | 88.75  | 88.75 |

### 1.2 結果 (a) の考察 :

結果の表 1 から、ユークリッド距離と類似度は同じ認識率を示した。レポート 1 からわかったように、二次元データにおいて各方法には優劣は見られたもののその差は微小なものだった。これらの方法を 256 次元に拡張したら、各次元にそれぞれの方法の差が平均化され、どんな方法であっても一定の認識率を予想できる。基本的な方法であるユークリッド距離と類似度はこうした高次元のデータ判別においても安定した認識率を果しており、識別問題において有用であることがわかった。

### 1.3 重み付きユークリッド距離法による認識結果 :

重みをコード作成便利のため、1, 2, ..., 256 と各次元にかかる重みが線形的に増加するものとし、重み付きユークリッド距離法を用いて文字画像認識した結果を下表 2 に示した。

表 2 重み付きユークリッド距離法による正誤表及び認識率

| データ番号 | 重み付きユークリッド |
|-------|------------|
| 0     | True       |
| 1     | True       |
| 2     | True       |
| 3     | True       |
| ...   | ...        |
| 77    | False      |
| 78    | True       |
| 79    | False      |
| 認識率 % | 81.25      |

表2から後方にあるバイトが重くカウントされる重みの配分で認識させると認識率が下がった。

## 2 判別分析・主成分分析

主成分分析法で判別評価基準にユークリッド距離法を用いて判別を行った結果一部を表3に示す。

表3 主成分分析による正誤表及び認識率

| データ番号 | 重み付きユークリッド |
|-------|------------|
| 0     | True       |
| 1     | True       |
| 2     | True       |
| 3     | True       |
| 4     | True       |
| 5     | True       |
| 6     | True       |
| 7     | True       |
| 8     | True       |
| 9     | True       |
| ..    | ...        |
| 70    | False      |
| 71    | True       |
| 72    | False      |
| 73    | True       |
| 74    | True       |
| 75    | True       |
| 76    | True       |
| 77    | True       |
| 78    | True       |
| 79    | False      |
| 認識率 % | 85.00      |

い認識率が見られたものの、ノイズに弱い欠点がある。主成分分析は多次元のデータから分析すべく主成分軸を探り出し、認識率を高める方法であり、特に画像のような圧縮性を持つ多次元データに対して優れているアプローチ方法である。

### 3.2 (b) 本講義について

本講義では内容の難易度はもともと高めである上、教員による解説時間が短いため習得するに当たって大量の時間を費やさなければならず、高い壁を感じた。ただし、パターン認識は人工知能がの発展に伴い、今後ますます多用されていくと考えられ習得するのに値する価値はあると思う。

## 3 パターン認識論について：

### 3.1 (a) 今までの判別方法の特徴

ユークリッド距離法と類似度は一貫して安定な認識率を示しており、判別方法としては汎用性のあるものであると考えられる。Nearest neighbor 法はデータに明確な境界線が存在していればもっとも高