

パターン認識・課題2

課題2.1 最近傍決定則による認識プログラムの作成

課題2.1.1 プログラムの作成

演習課題1で行った最近傍法を行うプログラムを作成せよ。

ただし、プログラムは以下の条件を満たすこと。

- 入力するパターンは、プログラム中に書くのではなく 外部ファイルから読み込むようにする。
- 今回は $5 \times 5 = 25$ 次元のデータを扱ったが、今後、次元数が変化することを考慮に入れる。
- 学習用データの数も5個 \times 3クラスとは限らないことを考慮に入れる。

課題2.1.2 上記プログラムを用いた数字の判別実験

課題1 (exercise1.html)の未知パターンを適用し、未知パターンがどの数字かを判別せよ。

課題2.2 k近傍決定則による認識プログラムの作成

課題2.2.1 プログラムの作成

課題2.2で行った最近傍法を行うプログラムを拡張し、k近傍法を実現せよ。

ただし、プログラムは以下の条件を満たすこと。

- kの値を引数で指定できるようにすること。

課題2.2.2 上記プログラムを用いた数字の判別実験

課題2.1.2と同様

課題2.3 プロトタイプの最近傍決定則による認識プログラムの作成

課題2.3.1 プログラムの作成

課題1で行ったプロトタイプの最近傍法を行うプログラムを作成せよ。

ただし、プログラムは以下の条件を満たすこと。

- プログラムは、
 - プロトタイプの作成
 - 認識をそれぞれ別のプログラムとして作成する。
- プロトタイプパターンは一旦ファイルに出力する。

課題2.3.2 上記プログラムを用いた数字の判別実験

Yasutomo KAWANISHI
課題2.1.2と同様

課題2.4 3つの手法の比較

課題2.1-2.3で実装した3つの手法について、外れ値への頑健性、クラス数や学習パターン数が増えたときの計算速度などの観点から利点・欠点を比較し、考察せよ。

課題2.5（調査課題）パターン認識の利用例

世の中で実際に利用されているパターン認識の例を挙げ、どのような仕組みで行われているかを述べよ。

たとえば、郵便局では7桁の郵便番号を自動的に読み込み、送り先を分類している。

ただし、文字認識以外の例をあげること。

参考文献を引用しつつ説明すること（引用箇所は明記すること）。

プログラム作成のヒント（課題2.1）

プログラムの流れ

1. 学習パターンを用意する。pattern2-0.dat, pattern2-1.dat,...,pattern9-4.dat
2. 未知パターンを読み込む。ファイル名はプログラムの引数で与えられるのが望ましい。
3. 未知パターンと学習パターンとの距離計算を行う。
4. 未知パターンがどの数字かを表示する。

データ形式(例)

各学習パターン、未知パターンは以下のフォーマットで与えられるものとする

データのフォーマット

```
5 5
0 1 1 1 0
0 0 0 0 1
0 0 0 1 0
0 0 1 0 0
1 1 1 1 0
```

学習パターンの読み込み例:読み込みファイル一覧をファイルに保存しておく

あるファイルに、読み込むべきファイル名を羅列しておいて、そのファイルを読み込む。

```
files.dat
p1.dat
p2.dat
...
p3.dat
```

```
char *listfile = argv[1];
FILE *files = fopen(listfile, "r"); //ファイル名保存ファイルを開く
char fileName[256];
for(i = 0; i < DATA_NUM; i++){
    fscanf(files, "%s", fileName); //ファイル名を読み込む
    FILE *dataFile = fopen(fileName, "r");
    /*
        データ読み込み
    */
}
```

距離計算例

多次元配列を2つ引数に取り, スカラー値を返す関数を作成する.

作成する関数は, $c = \text{distance}(a, b)$ の演算であり, a, b を引数にとって c を return するものとする.

```
double data_dist(MojiData *mdata_1, MojiData *mdata_2){
    double result = 0.0;

    /* mdata_1とmdata_2のユークリッド距離を計算する */

    return result;
}
```

プログラム作成のヒント(課題2.3)

プログラムの流れ

プロトタイプ作成

- 1つのクラスの学習パターンを用意する. pattern2-0.dat, pattern2-1.dat,...,pattern2-4.dat
- 学習パターンからプロトタイプを作成する. prototype2.dat, prototype7.dat, prototype9.dat
- プロトタイプを保存する.

認識処理

1. 各プロトタイプを読み込む
2. 未知パターンを読み込む. ファイル名はプログラムの引数で与えられるのが望ましい.
3. 未知パターンと各プロトタイプとの距離計算を行う.
4. 未知パターンがどの数字かを表示する.

割り算例

多次元配列をスカラー値で割る関数を作成する.

作成する関数は, $c = a/b$ の演算であり, a, b, c を引数にとるものとする.

```
void data_div(MojiData *mdata_1, double val, MojiData *mdata_out){
    int i, j;
    for(i = 0; i < mdata_1->h; i++){
        for(j = 0; j < mdata_1->w; j++){
            /* mdata_1の要素をvalで割ってmdata_outに格納する */

        }
    }
}
```

ファイルへの書き込み例

```
File* file = fopen("ファイル名", "w");
int input = 5;
fprintf(file, "%d", input);
fclose(file);
```

必ずfopenしたらfcloseすること.

構造体をファイルに保存する関数を作成しておくとう便利.

レポートの例(課題2.1)

最近傍決定則によって認識した結果, 各未知パターンに対し, 最も距離が近い学習パターンとその距離は表1のようになった.

従って, 〇%が正しく判断された.
ただし, △△については正しく判断することが出来なかった.
その理由は考察で述べる.

注意:数字は適当

レポートの例(課題2.2)

k近傍決定則によって認識した結果, 各未知パターンに対し, 最も距離が近いk個の学習パターンとその距離は表1のようになった.

k個の近傍の中で多数決をとった結果も表1に載せる.

従って, 〇%が正しく判断された.
ただし, △△については正しく判断することが出来なかった.
その理由は考察で述べる.

レポートの例(課題2.3)

(省略)同様に書くこと.

提出期限

2018/5/14

(レポートの内容や提出時の注意事項は, [こちら](#) (../report.html))

[戻る](#)