

第 11 章

第 9 章 評価法とチューニング 演習問題

11.1 演習の目的

Scilab を用いて特徴量を評価し、より識別能力の高い特徴空間を構成する方法を学びます。

11.2 準備

実践第 5 章で用いた数字画像データを用いて、抽出した平均・分散に加えて、以下に示す歪度・尖度（それぞれ x, y 方向）を特徴として計算します。歪度（わいど）は 3 次モーメントとも呼ばれ、分布の左右非対称性を表します。尖度（せんど）は 4 次モーメントとも呼ばれ、分布の尖り具合、あるいは分布の裾の重さを表しています*1。

歪度

$$\begin{aligned} S_X &= \frac{1}{\sigma_X^3} \sum_{x=1}^X \sum_{y=1}^Y (x - \mu_X)^3 \hat{I}_{x,y} \\ S_Y &= \frac{1}{\sigma_Y^3} \sum_{x=1}^X \sum_{y=1}^Y (y - \mu_Y)^3 \hat{I}_{x,y} \end{aligned} \quad (11.1)$$

尖度

$$\begin{aligned} F_X &= \frac{1}{\sigma_X^4} \sum_{x=1}^X \sum_{y=1}^Y (x - \mu_X)^4 \hat{I}_{x,y} \\ F_Y &= \frac{1}{\sigma_Y^4} \sum_{x=1}^X \sum_{y=1}^Y (y - \mu_Y)^4 \hat{I}_{x,y} \end{aligned} \quad (11.2)$$

実践演習 11-1

数字画像データの x, y 方向それぞれの平均・分散・歪度・尖度の計 8 次元から、識別に有効な 2 次元を選択して特徴空間を構成したい。定量的評価として、すべての 2 次元の組み合わせについてクラス内分散・クラス間分散比を計算せよ。

*1 高校数学の美しい物語「歪度、尖度の定義と意味」<http://mathtrain.jp/waidosendo> 参照

```

clear
DATADIR = 'number\';
n = 100; // データ数
c = 10; // クラス数
d = 8; // 特徴ベクトルの次元数
y = matrix(repmat([0:9],[10,1]),[100,1]); // 正解ベクトル

// ファイルを読み込み、特徴量を計算
function result = feature(filename)
    im = im2double(imread(DATADIR+filename));
    [h w] = size(im);
    //反転
    im = 1 - im;
    //正規化
    im = im ./ sum(im);
    //平均
    mx = sum(im,'r') * (1:w)';
    my = sum(im,'c')' * (1:h)';
    //分散
    vx = sum(im,'r') * (((1:w) - mx) .^2)';
    vy = sum(im,'c')' * (((1:h) - my) .^2)';
    //歪度
    sx = sum(im,'r') * (((1:w) - mx) .^3)' / sqrt(vx)^3;
    sy = sum(im,'c')' * (((1:h) - my) .^3)' / sqrt(vy)^3;
    //尖度
    fx = sum(im,'r') * (((1:w) - mx) .^4)' / vx^2;
    fy = sum(im,'c')' * (((1:h) - my) .^4)' / vy^2;
    result = [mx, my, vx, vy, sx, sy, fx, fy];
endfunction

// クラス内分散・クラス間分散比の計算
function[result] = wcbc(va, vb)
    ...
endfunction

// Mainプログラム
// 読み込み・特徴抽出
X=[];
for i=0:9 do
    for j=0:9 do
        X = [X; feature('number'+string(i)+'_'+string(j)+'.pgm')];
    end
end

//標準化
m = mean(X,'r');
s = stdev(X, 'r');
normX = (X - repmat(m,[n,1])) ./ repmat(s,[n,1]);
csvWrite([normX, y], 'number8.csv');

// 定量的評価
J = zeros(d,d);
for i = 1:d-1
    for j = i+1:d
        J(i, j) = wcbc(normX(:,i), normX(:,j));
    end
end
end

```

実践演習 11-2

実践演習 11-1 の結果から上位 3 つ程度の組み合わせを選択し、plot2d 関数を用いた散布図表示による定性的評価を行って、もっとも良いと思われる 2 次元特徴量を求めよ。

実践演習 11-3

特徴空間の評価法をクラス内分散・クラス間分散比から 1-NN 法の識別率に置き換え、結果の違いを確認せよ。