

第13回講義「統計学2」

カーネル密度推定法(2)

[1] 尤度交差確認法による

カーネル関数バンド幅の決定

[2] カーネル密度推定法による

手書き数字の認識

尤度交差確認法(1)

限られた訓練標本を用いて、できるだけ汎化性能のある確率密度関数推定モデルを選択する有力な手法として、**尤度交差確認法**がある。

尤度交差確認法は、訓練標本を「推定用」と「確認用」に分割して、推定用の標本で確率密度関数を推定し、確認用の標本でその尤度を評価する。

その際、訓練標本の分割の仕方にモデル選択の結果が大きく依存することが問題となる。

尤度交差確認法(2)

この依存性を軽減するため、次の方法を取る。

1. 訓練標本を h 個の部分集合に等分割する。
2. $(h-1)$ 個の部分集合を「推定用」に、残りの1個の部分集合を「確認用」にして、尤度を計算する。
3. $(h-1)$ 個の部分集合の選び方は h 通りあるため、 h 回の尤度の評価を行う。
4. h 個の尤度の平均値を最終的な尤度の評価値とする。通常、 $h = 5$ または $h = 10$ を用いる。

カーネル関数バンド幅の決定

1. バンド幅の候補をいくつか用意する: $\{b_j\}_{j=1}^J$
2. 訓練標本 $\{x_i\}_{i=1}^n$ を h 個の部分集合に分割する。
3. 各モデル b_j に対して、尤度交差確認法により対数尤度の平均評価値を求める。
4. 対数尤度の平均評価値が最大となる b_{j^*} を選ぶ。
5. 選ばれたモデル b_{j^*} に対して、全ての訓練標本を用いて、確率密度関数の推定 $\hat{p}(x)$ を求める。

練習問題18

カーネル密度推定法を用いて、先に配布した‘0’から‘9’の手書き数字データの認識を行いなさい。ただし、訓練データ X を $h = 5$ として分割しなさい。また、ガウスカーネルのバンド幅として、 $\{ 0.1, 0.2, 0.3, 0.4, 0.5 \}$ を試してみなさい。最適なバンド幅が決定されたら、訓練データ全てを用いて各カテゴリの確率密度関数を推定し、テストデータ T の認識実験を行いなさい。

練習問題18(続き)

ヒント: 以下のプログラムを参考にしなさい。

```
clear all
load digit.mat X T
[d, nx, nc] = size(X);

% ガウスカーネル関数のバンド幅
b = [0.1:0.1:0.5];

% 尤度交差確認法
% 訓練データの分割数
h = 5;
sn = nx/h;

lpd = zeros(1, length(b));
```

練習問題18(続き)

```
for s = 1:length(b)
  for k = 1:h
    for c = 1:nc
      for i = 1:nx
        m = mod(i, sn) + 1;
        if m != k
          continue
        end
        xx = X(:, i, c); kpdf = 0;
        for j = 1:nx
          m = mod(j, sn) + 1;
          if m == k
            continue
          end
          x = X(:, j, c); kpdf += exp(-(xx-x)'*(xx-x)/(2*b(s)^2));
        end
        lpd(s) += log(kpdf) - d*log(b(s));
      end
    end
  end
end
```

練習問題18(続き)

```
end
end
end
end

% 最適なバンド幅の決定
[vmax pmax] = max(lpd); bop = b(pmax);

% テストデータの認識
[d, nt, nc] = size(T);

% 各カテゴリの尤度の計算
for ct = 1:nc
    for i = 1:nt
        xx = T(:, i, ct);
        for c = 1:nc
            kpdf = 0;
```

練習問題18(続き)

```
for j = 1:nx
    x = X(:, j, c);
    kpdf += exp(-(xx-x)'*(xx-x)/(2*bop^2));
end
p(ct, i, c) = log(kpdf);
end
end
end

% 混同行列の作成
[pmax P] = max(p, [], 3);
for ct = 1:nc
    for c = 1:nc
        C(ct, c) = sum(P(ct, :) == c);
    end
end
end
```