パターン認識
Pattern Recognition

佐藤真一
Shin'ichi Satoh

国立情報学研究所
National Institute of Informatics

May 9, 2023

k-近傍識別器

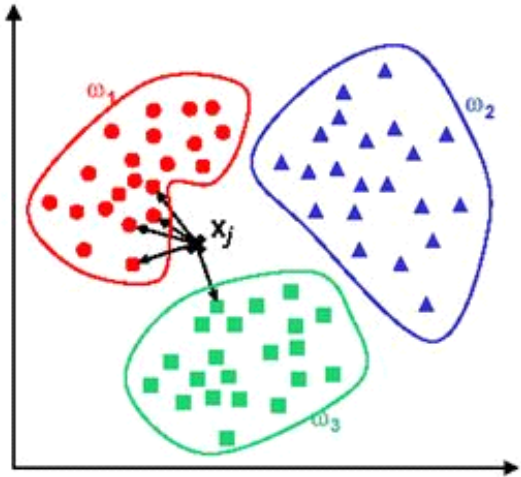
今、c クラスを考え、各クラス ω_i は学習データ中 n_i 点を有し $n_1 + n_2 + \cdots + n_c = n$ が成り立つとすると、

$$P(\omega_i|\mathbf{x}) = \frac{p_n(\mathbf{x}|\omega_i)P(\omega_i)}{p_n(\mathbf{x})}.$$

ある点 x が与えられたときに、k_n 個の最近傍点を選び出すこととする。
k_n 個中 k_i 個の点がクラス ω_i に所属するとすると、

$$p_n(x|\omega_i) = \frac{k_i}{n_i V_n}.$$

k-近傍識別器



事前確率は以下のように考えることができる

$$P(\omega_i) = \frac{n_i}{n}.$$

ベイズ識別則により、事後確率は以下の通り計算できる

$$P(\omega_i|x) = \frac{p_n(x|\omega_i)P(\omega_i)}{p_n(x)} = \frac{\frac{k_i}{n_i V_n}\frac{n_i}{n}}{\frac{k_n}{nV_n}}$$

$$= \frac{k_i}{k_n}$$

ただし $p_n(x) = \frac{k_n}{nV_n}$

k-近傍識別器

k-近傍識別則:

> 点 x が与えられたとき、それを中心としてちょうど k 個の点を含む超球を考えた時に、x のクラスとしてその超球の中で最多数を占めるクラスとする

$$P(\omega_i|x) = \frac{p_n(x|\omega_i)P(\omega_i)}{p_n(x)} = \frac{k_i}{k_n}$$

$k = 1$ の時、最近傍識別則、あるいは最近傍識別器と呼ぶ

最近傍識別則の誤り率

以下を示したい:

$$P^* \leq P \leq P^*(2 - \frac{c}{c-1}P^*) \leq 2P^*.$$

ただし P は無限の学習データがある場合の最近傍識別則による誤り率、P^* は誤り率の理論限界 (ベイズ誤り率)

最近傍識別則の誤り率

n 個のラベル付きプロトタイプ (学習データ): $\mathcal{D}^n = \{x_1, \cdots, x_n\}$.
テストしたい点 x に最も近いプロトタイプ: $x' \in \mathcal{D}^n$.
最近傍識別則: x を最も近いプロトタイプ x' のラベルに識別
x' のラベルに対応する確率変数 : θ'
$\theta' = \omega_i$ となる確率 : 事後確率 $P(\omega_i|x')$
サンプル数が十分多いとすると、x' が十分 x に近いと仮定して差し支えないので、
$P(\omega_i|x') \simeq P(\omega_i|x)$
従って、最近傍識別則は真の事後確率とほぼ一致する

最近傍識別則の誤り率

以下の通り定義すると、

$$P(\omega_m|x) = max_i P(\omega_i|x).$$

ベイズ決定測は常に ω_m を選ぶ
学習データが無限にあった場合の条件付平均誤り率を $P(e|x)$、条件なし平均誤り率を $P(e)$ とすると、

$$P(e) = \int P(e|x)p(x)dx.$$

さらに $P^*(e|x)$ を $P(e|x)$ の可能な最小値、P^* を $P(e)$ の可能な最小値とすると、

$$P^*(e|x) = 1 - P(\omega_m|x) \text{ 及び } P^* = \int P^*(e|x)p(x)dx.$$

最近傍識別則の誤り率

$P_n(e)$ を学習サンプルが n 個の場合の誤り率とし、もし

$$P = \lim_{n \to \infty} P_n(e)$$

ならば、以下を示したい:

$$P^* \leq P \leq P^*(2 - \frac{c}{c-1} P^*)$$

x' を x の最近傍とすると、

$$P(e|x) = \int P(e|x, x') p(x'|x) dx'$$

条件付確率密度 $p(x'|x)$ の算出は容易ではない

最近傍識別則の誤り率

しかし、x' は x の最近傍なので、$p(x'|x)$ は x を中心としたデルタ関数に漸近すると考えられる
今、x を中心とした超球 $\mathcal{S}$ 内にサンプルが存在する確率を考えると、

$$P_{\mathcal{S}} = \int_{x' \in \mathcal{S}} p(x')dx'$$

全 n サンプルが $\mathcal{S}$ 外に落ちる確率は $(1 - P_{\mathcal{S}})^n$ であり、これは n が無限に近づくにしたがって 0 に漸近する
従って、x' は x に確率的に収束し、期待通り $p(x'|x)$ はデルタ関数に収束する

最近傍識別則の誤り率

条件付誤り率 $P_n(e|x, x')$ について考える

- x'_n: サンプル数 n の中の x の最近傍
- n 個の独立に抽出されたラベル付きサンプル $(x_1, \theta_1), \ldots, (x_n, \theta_n)$

これらの対は以下によりそれぞれ独立に生成されたと仮定する:

1. θ_j に対応する状態 ω_j を確率 $P(\omega_j)$ で選ぶ
2. 次いで x_j を確率分布 $p(x|\omega_j)$ に従って選ぶ

最近傍識別則の誤り率

識別の際、実際は x のラベルは θ であり、かつ x'_n (ラベル θ'_n) が x の最近傍だと仮定する x'_n を選んだ時の状態と x を選んだ時の状態とは独立であるため、

$$P(\theta, \theta'_n | x, x'_n) = P(\theta|x)P(\theta'_n|x'_n).$$

この場合の条件付誤り率は

$$\begin{aligned} P(e|x, x'_n) &= 1 - \sum_{i=1}^{c} P(\theta = \omega_i, \theta'_n = \omega_i | x, x'_n) \\ &= 1 - \sum_{i=1}^{c} P(\omega_i|x)P(\omega_i|x'_n). \end{aligned}$$

最近傍識別則の誤り率

これまでの議論により

$$P(e|x) = \int P(e|x, x')p(x'|x)dx'$$
$$p(x'|x) = \delta(x' - x)$$

さらに

$$\lim_{n\to\infty} P_n(e|x) = \int [1 - \sum_{i=1}^{c} P(\omega_i|x)P(\omega_i|x'_n)]\delta(x'_n - x)dx'_n$$
$$= 1 - \sum_{i=1}^{c} P^2(\omega_i|x).$$

最近傍識別則の誤り率

最近傍識別器の誤り率は以下に漸近する

$$
\begin{aligned}
P &= \lim_{n\to\infty} P_n(e) \\
&= \lim_{n\to\infty} \int P_n(e|x)p(x)dx \\
&= \int [1 - \sum_{i=1}^{c} P^2(\omega_i|x)]p(x)dx.
\end{aligned}
$$

最近傍識別則の誤り率

以下の定義を思い起こそう

$$P^*(e|x) = 1 - P(\omega_m|x).$$

$P(\omega_m|x)$ が決まっている状況、すなわち P^* が決まっている状況で、$\sum_{i=1}^{c} P^2(\omega_i|x)$ がどれだけ小さくなりうるか知りたい

最近傍識別則の誤り率

以下のように書ける

$$\sum_{i=1}^{c} P^2(\omega_i|x) = P^2(\omega_m|x) + \sum_{i \neq m} P^2(\omega_i|x).$$

以下の条件の時に上記を最小化したい

- $P(\omega_i|x) \geq 0$
- $\sum_{i \neq m} P(\omega_i|x) = 1 - P(\omega_m|x) = P^*(e|x)$.

最近傍識別則の誤り率

$P(\omega_m|x)$ 以外のすべての事後確率を同じ値とすることにより、$\sum_{i=1}^{c} P^2(\omega_i|x)$ は最小となる

$$P(\omega_i|x) = \begin{cases} \dfrac{P^*(e|x)}{c-1} & i \neq m \\ 1 - P^*(e|x) & i = m \end{cases}$$

従って

$$\sum_{i=1}^{c} P^2(\omega_i|x) \geq (1 - P^*(e|x))^2 + \frac{(P^*(e|x))^2}{c-1}$$

$$1 - \sum_{i=1}^{c} P^2(\omega_i|x) \leq 2P^*(e|x) - \frac{c}{c-1}(P^*(e|x))^2.$$

これにより直ちに $P \leq 2P^*$

最近傍識別則の誤り率

よりタイトな上界のため

$$
\begin{aligned}
Var[P^*(e|x)] &= \int [P^*(e|x) - P^*]^2 p(x) dx \\
&= \int (P^*(e|x))^2 p(x) dx - P^{*2} \geq 0
\end{aligned}
$$

従って

$$
\int (P^*(e|x))^2 p(x) dx \geq (P^*)^2
$$

等号は $P^*(e|x)$ の分散が 0 の時に限り成り立つ

従って

$$P^* \leq P \leq P^*(2 - \frac{c}{c-1}P^*)$$

最近傍識別則の誤り率

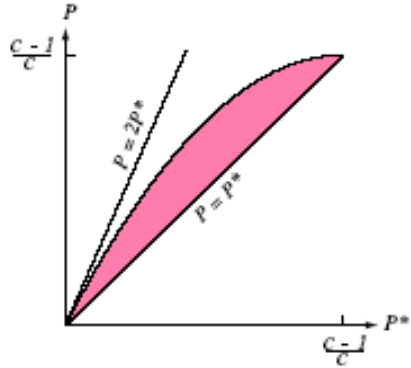


FIGURE 4.14. Bounds on the nearest-neighbor error rate P in a c-category problem given infinite training data, where P^* is the Bayes error (Eq. 52). At low error rates, the nearest-neighbor error rate is bounded above by twice the Bayes rate. From: Richard O. Duda, Peter E. Hart, and David G. Stork, *Pattern Classification*. Copyright © 2001 by John Wiley & Sons, Inc.

最近傍識別則の誤り率

- サンプル数が有限の場合、最近傍識別則はどの程度うまく働くか
- 精度はどのくらい早く漸近値に収束するか
- 実際には、収束は任意に遅く、誤り率 $P_n(e)$ は n に従って単調に減少するとも限らない
- 実際の確率分布の構造を仮定しなければ、漸近的結果以上を得るのは困難

k-近傍識別則

最近傍則を拡張し、k 個の近傍を得て、そのラベルに基づいて投票することにより決定することができる
二つの場合について考えられる

- k が奇数: 引き分けなし
- k が偶数: 引き分けを reject

k-近傍則の誤り率 P_{kNN} は:

$$\frac{1}{2}P^* \leq P_{2NN} \leq P_{4NN} \leq \cdots \leq P^* \leq \cdots \leq P_{3NN} \leq P_{NN} \leq 2P^*.$$

興味のある人は Fukunaga “Statistical Pattern Recognition” の第 7 章を参照のこと

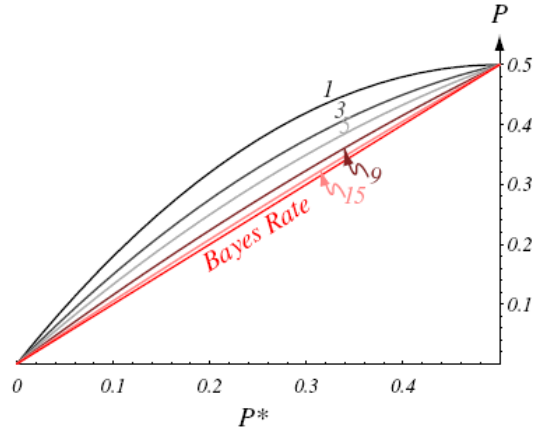


FIGURE 4.16. The error rate for the k-nearest-neighbor rule for a two-category problem is bounded by $C_k(P^*)$ in Eq. 54. Each curve is labeled by k; when $k = \infty$, the estimated probabilities match the true probabilities and thus the error rate is equal to the Bayes

演習

二つのクラスの条件付確率分布として、二つの正規分布を考える

Class 1 $\Sigma_1 = I,\ \mu_1 = [s, 0, \cdots, 0]^T$

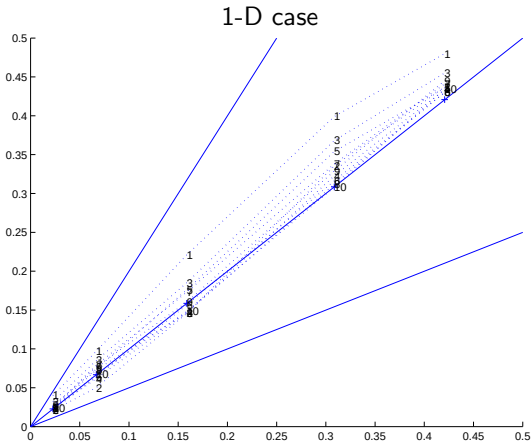
Class 2 $\Sigma_2 = I,\ \mu_2 = [-s, 0, \cdots, 0]^T$

ベイズ誤り率は

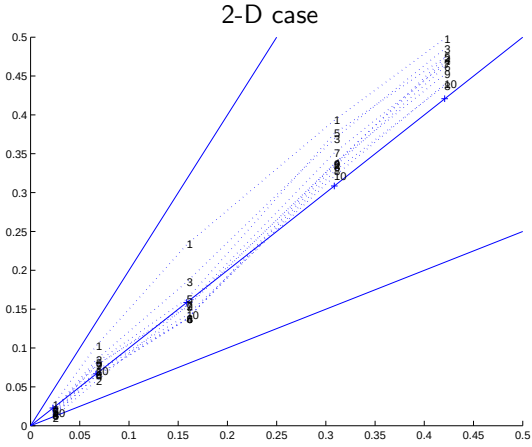
$$
\begin{aligned}
P^* &= \int_s^\infty \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty N(0, I) dx_1 \cdots dx_n \\
&= \frac{1}{2}(1 - \mathrm{erf}(\frac{s}{\sqrt{2}})) \\
\mathrm{erf}(x) &= \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt
\end{aligned}
$$

クラス 1 と 2 からそれぞれ n サンプル得たとして、k-近傍識別則の誤り率を計測してみよう

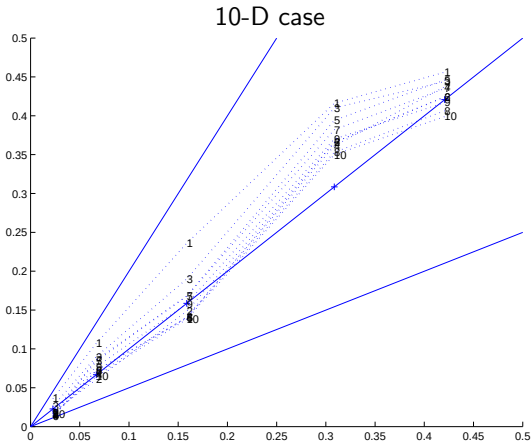
Exercise (knnclass.py)



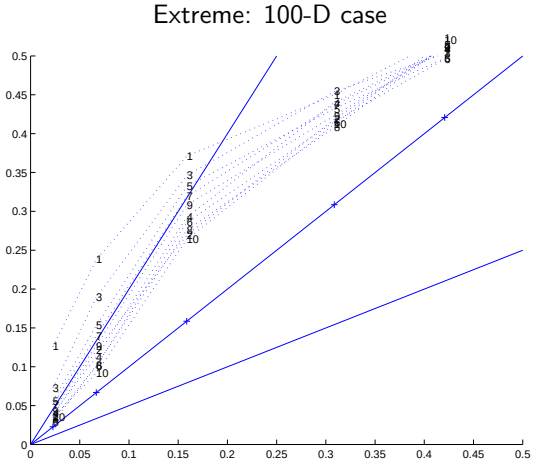
Exercise


2-D case
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5

Exercise


10-D case
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5

Exercise


Extreme: 100-D case
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5

演習

MNIST データに対して k-近傍識別器を適用してみよ
mntest.py を少し修正するだけで対応できるはず
強くお勧め (しかし宿題ではない)

最近傍識別を実装してみたところ、97%の精度を達成
k-近傍識別器ではどうなるか？