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 16, 2023

宿題について

- 5 月 9–10 日の ITC LMS のシステム障害のため、5 月 9 日締め切りであった 4 月 25 日出題の宿題の締め切りを一週間延長し 5 月 16 日とすることとした。
- よって直近の宿題の締め切りは以下の通り
  - 4 月 25 日出題分、締め切り 5 月 9 日 → 5 月 16 日に延長
  - 5 月 2 日出題分、締め切り 5 月 16 日
  - 5 月 9 日宿題なし
  - 5 月 16 日 (本日) 出題分、締め切り 5 月 30 日
- PDF フォーマットにて ITC LMS に提出すること。
- Jupyter Notebook (ipynb) や Python (py) 形式の提出はおやめください。

今日の予定

- サンプルからの誤り率推定
- 再代入法 (R 法)
- 分割学習法 (H 法)
- 交差確認法 (CV 法)
- ブートストラップ法

識別誤り率推定

簡単のため、2 クラス分類について考える。識別器は以下の関数により定義できる

$$h(X) \underset{\omega_2}{\overset{\omega_1}{\lessgtr}} 0$$

ただし $h(X)$ はベクトル X の識別関数。この識別器の識別誤り率は

$$\varepsilon_1 = \int_{h(X)>0} p_1(X)dX = \int u(h(X))p_1(X)dX$$

ただし $u(\cdot)$ はステップ関数、$p_i(X)$ はクラス i の確率密度関数。

いまフーリエ変換とフーリエ逆変換を考える。

$$\mathscr{F}[x(t)] = X(\omega) = \int_{-\infty}^{\infty} x(t)\mathrm{e}^{-j\omega t}dt$$

$$\mathscr{F}^{-1}[X(\omega)] = x(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} X(\omega)\mathrm{e}^{j\omega t}d\omega$$

ステップ関数 $u(t)$ は

$$u(t) = \begin{cases} 0 & t < 0 \\ \frac{1}{2} & t = 0 \\ 1 & t > 0 \end{cases}$$

そのフーリエ変換

$$\begin{aligned}\mathscr{F}[u(t)] &= \frac{1}{2}\mathscr{F}[1] + \int_0^\infty \mathrm{e}^{-j\omega t} dt \\ &= \frac{1}{2}(2\pi\delta(\omega)) - \frac{1}{j\omega}\left[\mathrm{e}^{-j\omega t}\right]_0^\infty \\ &= \pi\delta(\omega) + \frac{1}{j\omega}\end{aligned}$$

そのフーリエ逆変換

$$u(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}[\pi\delta(\omega) + \frac{1}{j\omega}]\mathrm{e}^{j\omega t} d\omega$$

よって

$$\begin{aligned}\varepsilon_1 &= \int_{h(X)>0} p_1(X)dX = \int u(h(X))p_1(X)dX \\ &= \frac{1}{2\pi}\iint [\pi\delta(\omega) + \frac{1}{j\omega}]\mathrm{e}^{j\omega h(X)}p_1(X)d\omega dX \\ &= \frac{1}{2} + \frac{1}{2\pi}\iint \frac{\mathrm{e}^{j\omega h(X)}}{j\omega}p_1(X)d\omega dX\end{aligned}$$

同様に

$$\varepsilon_2 = \int_{h(X)<0} p_2(X)dX = \frac{1}{2} - \frac{1}{2\pi}\iint \frac{e^{j\omega h(X)}}{j\omega} p_2(X)d\omega dX.$$

全体の識別誤り率は

$$\begin{aligned}\varepsilon =& P_1\varepsilon_1 + P_2\varepsilon_2 \\ =& \frac{1}{2} + \frac{1}{2\pi}\iint \frac{e^{j\omega h(X)}}{j\omega}\tilde{p}(X)d\omega dX\end{aligned}$$

ただし

$$\tilde{p}(X) = P_1 p_1(X) - P_2 p_2(X).$$

標本からの誤り率推定

有限の標本 (サンプル) のみ利用可能な場合の「誤り数計数法」の正当性について考える。
その場合、$p_i(X)$ は以下により置き換えられる

$$\hat{p}_i(X) = \frac{1}{N_i} \sum_{j=1}^{N_i} \delta(X - \mathbf{X}_j^{(i)})$$

ただし $\mathbf{X}_j^{(i)}$ は $p_i(X)$ から抽出した N_i 個の評価用の標本 (test sample) である。

そうすると誤り率の推定は

$$\begin{aligned}\hat{\varepsilon} =& \frac{1}{2} + \frac{1}{2\pi} \iint \frac{\mathrm{e}^{j\omega h(X)}}{j\omega} [\frac{P_1}{N_1} \sum_{j=1}^{N_1} \delta(X - \mathbf{X}_j^{(1)}) - \frac{P_2}{N_2} \sum_{j=1}^{N_2} \delta(X - \mathbf{X}_j^{(2)})] d\omega dX \\ =& \frac{1}{2} + \frac{P_1}{N_1} \sum_{j=1}^{N_1} \boldsymbol{\alpha}_j^{(1)} - \frac{P_2}{N_2} \sum_{j=1}^{N_2} \boldsymbol{\alpha}_j^{(2)}\end{aligned}$$

ただし

$$\boldsymbol{\alpha}_j^{(i)} = \frac{1}{2\pi} \int \frac{\mathrm{e}^{j\omega h(\mathbf{X}_j^{(i)})}}{j\omega} d\omega = \frac{sign(h(\mathbf{X}_j^{(i)}))}{2}$$

これにより

$$\begin{aligned}\frac{1}{N_1}\sum_{j=1}^{N_1}\boldsymbol{\alpha}_j^{(1)} &= \frac{1}{2N_1}[(\#\ \omega_1\text{-errors}) - (\#\ \omega_1\text{-corrects})] \\ &= \frac{1}{N_1}(\#\ \omega_1\text{-errors}) - \frac{1}{2}.\end{aligned}$$

同様に

$$\frac{1}{N_2}\sum_{j=1}^{N_2}\boldsymbol{\alpha}_j^{(2)} = -\frac{1}{N_2}(\#\ \omega_2\text{-errors}) + \frac{1}{2}.$$

まとめると

$$\hat{\varepsilon} = P_1\frac{(\#\ \omega_1\text{-errors})}{N_1} + P_2\frac{(\#\ \omega_2\text{-errors})}{N_2}.$$

推定誤り率の平均と分散

推定誤りの評価用標本に対する平均と分散を求めよう。

$$
\begin{aligned}
E_t\{\boldsymbol{\alpha}_j^{(i)}\} =& \bar{\alpha}_i = \frac{1}{2\pi}\iint \frac{\mathrm{e}^{j\omega h(X)}}{j\omega} p_i(X) d\omega dX \\
=& \begin{cases} \varepsilon_1 - \dfrac{1}{2} & (i=1) \\ \dfrac{1}{2} - \varepsilon_2 & (i=2) \end{cases} \\
E_t\{{\boldsymbol{\alpha}_i^{(i)}}^2\} =& E_t\{[\frac{1}{2\pi}\int \frac{\mathrm{e}^{j\omega h(\mathbf{X})}}{j\omega} d\omega]^2\} = E_t\{[\frac{1}{2} sign(h(\mathbf{X}))]^2\} \\
=& \frac{1}{4} \\
E_t\{\boldsymbol{\alpha}_j^{(i)}\boldsymbol{\alpha}_\ell^{(k)}\} =& \bar{\alpha}_i\bar{\alpha}_k \quad (i \neq k \text{ or } j \neq \ell)
\end{aligned}
$$

推定誤り率の平均と分散

まとめると

$$
\begin{aligned}
E_t\{\hat{\varepsilon}\} =& \frac{1}{2} + P_1\bar{\alpha_1} - P_2\bar{\alpha_2} \\
=& \frac{1}{2} + P_1(\varepsilon_1 - \frac{1}{2}) - P_2(\frac{1}{2} - \varepsilon_2) = \varepsilon \\
Var_t\{\hat{\varepsilon}\} =& \frac{P_1^2}{N_1} Var_t\{\boldsymbol{\alpha}_j^{(1)}\} + \frac{P_2^2}{N_2} Var_t\{\boldsymbol{\alpha}_j^{(2)}\} \\
=& \frac{P_1^2}{N_1}[\frac{1}{4} - (\varepsilon_1 - \frac{1}{2})^2] + \frac{P_2^2}{N_2}[\frac{1}{4} - (\frac{1}{2} - \varepsilon_2)^2] \\
=& P_1^2 \frac{\varepsilon_1(1-\varepsilon_1)}{N_1} + P_2^2 \frac{\varepsilon_2(1-\varepsilon_2)}{N_2}.
\end{aligned}
$$

すなわち、$\hat{\varepsilon}$ は $h(X)$ によらず不偏 (unbiased) で一致性のある (consistent) 誤り率推定である。

別解

$\hat{\tau}_i$ をクラス i について識別を誤った標本数とする。確率変数 $\hat{\tau}_1$ と $\hat{\tau}_2$ は独立であり、二項分布に従うので、

$$\begin{aligned} Pr\{\hat{\tau}_1 = \tau_1, \hat{\tau}_2 = \tau_2\} &= \prod_{i=1}^{2} Pr\{\hat{\tau}_i = \tau_i\} \\ &= \prod_{i=1}^{2} \left(\begin{array}{c} N_i \\ \tau_i \end{array} \right) \varepsilon_i^{\tau_i} (1 - \varepsilon_i)^{N_i - \tau_i}. \end{aligned}$$

ω_i 誤り、すなわち ε_i は、$\dfrac{\hat{\tau}_i}{N_i}$ として推定できる

$$\hat{\varepsilon} = \sum_{i=1}^{2} P_i \frac{\hat{\tau}_i}{N_i}.$$

別解

二項分布の平均と分散から、

$$
\begin{aligned}
E\{\hat{\varepsilon}\} &= P_1\varepsilon_1 + P_2\varepsilon_2 = \varepsilon \\
Var\{\hat{\varepsilon}\} &= P_1^2\frac{\varepsilon_1(1-\varepsilon_1)}{N_1} + P_2^2\frac{\varepsilon_2(1-\varepsilon_2)}{N_2}
\end{aligned}
$$

ベイズ誤り率の上界と下界

もし有限個の標本しか利用できない場合、われわれはどれを学習に用いるか、どれを評価に用いるかに関し、妥協を強いられる。
ここで識別誤りを学習セット (design/training set) と評価セット (test set) という二つの集合の関数として考える

$$\varepsilon(\mathscr{P}_D, \mathscr{P}_T)$$

ただし $\mathscr{P}$ はクラスごとの分布を表す。例えば

$$\mathscr{P} = \{p_1(X), p_2(X)\}.$$

ベイズ誤り率の上界と下界

もし識別器が評価用のデータの分布に対するベイズ識別器ならば、識別誤りは最小となる

$$\varepsilon(\mathscr{P}_T, \mathscr{P}_T) \leq \varepsilon(\mathscr{P}_D, \mathscr{P}_T).$$

真の $\mathscr{P}$ に対するベイズ誤り率は $\varepsilon(\mathscr{P}, \mathscr{P})$ となる。真の $\mathscr{P}$ は得られないため、有限個の標本による推定 $\hat{\boldsymbol{\mathscr{P}}} = (\hat{\mathbf{p}}_1(X), \hat{\mathbf{p}}_2(X))$ に基づく $\varepsilon(\mathscr{P}, \mathscr{P})$ の上界と下界について考えよう。

$$\begin{aligned}\varepsilon(\mathscr{P}, \mathscr{P}) &\leq \varepsilon(\hat{\boldsymbol{\mathscr{P}}}, \mathscr{P}) \\ \varepsilon(\hat{\boldsymbol{\mathscr{P}}}, \hat{\boldsymbol{\mathscr{P}}}) &\leq \varepsilon(\mathscr{P}, \hat{\boldsymbol{\mathscr{P}}}).\end{aligned}$$

誤り数計数法は不偏で一致性があることから、

$$\varepsilon(\hat{\boldsymbol{\mathscr{P}}}, \mathscr{P}) = E_{\hat{\boldsymbol{\mathscr{P}}}_T}\{\varepsilon(\hat{\boldsymbol{\mathscr{P}}}, \hat{\boldsymbol{\mathscr{P}}}_T)\}$$

ただし $\hat{\boldsymbol{\mathscr{P}}}_T$ は $\hat{\boldsymbol{\mathscr{P}}}$ とは独立の別の集合。

ベイズ誤り率の上界と下界

再掲:

$$\begin{aligned}\varepsilon(\mathscr{P}, \mathscr{P}) \leq & \varepsilon(\hat{\boldsymbol{\mathscr{P}}}, \mathscr{P}) \\ \varepsilon(\hat{\boldsymbol{\mathscr{P}}}, \hat{\boldsymbol{\mathscr{P}}}) \leq & \varepsilon(\mathscr{P}, \hat{\boldsymbol{\mathscr{P}}}) \\ \varepsilon(\hat{\boldsymbol{\mathscr{P}}}, \mathscr{P}) = & E_{\hat{\boldsymbol{\mathscr{P}}}_T}\{\varepsilon(\hat{\boldsymbol{\mathscr{P}}}, \hat{\boldsymbol{\mathscr{P}}}_T)\}.\end{aligned}$$

同様に、

$$E\{\varepsilon(\mathscr{P}, \hat{\boldsymbol{\mathscr{P}}})\} = \varepsilon(\mathscr{P}, \mathscr{P}).$$

まとめると

$$E\{\varepsilon(\hat{\boldsymbol{\mathscr{P}}}, \hat{\boldsymbol{\mathscr{P}}})\} \leq \varepsilon(\mathscr{P}, \mathscr{P}) \leq E_{\hat{\boldsymbol{\mathscr{P}}}_T}\{\varepsilon(\hat{\boldsymbol{\mathscr{P}}}, \hat{\boldsymbol{\mathscr{P}}}_T)\}.$$

分割学習法 (Holdout/H 法)

$\mathscr{P}$ から独立した二つの標本集合 $\hat{\boldsymbol{\mathscr{P}}}$ と $\hat{\boldsymbol{\mathscr{P}}}_T$ を得れば、$\varepsilon(\hat{\boldsymbol{\mathscr{P}}}, \hat{\boldsymbol{\mathscr{P}}}_T)$ は $\hat{\boldsymbol{\mathscr{P}}}$ を学習セット、$\hat{\boldsymbol{\mathscr{P}}}_T$ を評価セットとして用いれば算出可能。
すでに見てきているように、$E_{\hat{\boldsymbol{\mathscr{P}}}_T}\{\varepsilon(\hat{\boldsymbol{\mathscr{P}}}, \hat{\boldsymbol{\mathscr{P}}}_T)\}$ はベイズ誤り率の上界である。
同様に $E_{\hat{\boldsymbol{\mathscr{P}}}}\{E_{\hat{\boldsymbol{\mathscr{P}}}_T}\{\varepsilon(\hat{\boldsymbol{\mathscr{P}}}, \hat{\boldsymbol{\mathscr{P}}}_T)\}\}$ も上界となる。
この手続きを分割学習法 (Holdout/H 法) と呼ぶ。

分割学習法 (Holdout/H 法)

1. 与えられたデータセット S を二つの互いに素の集合、学習セット S_D と評価セット S_T に分割する
2. S_D を用いて識別器を学習する
3. S_T を用いて誤り率を測定する

再代入法 (Resubstitution/R 法)

$E\{\varepsilon(\hat{\mathscr{P}}, \hat{\mathscr{P}})\}$ はベイズ誤り率の下界となる。
これは $\hat{\mathscr{P}}$ をベイズ識別器の学習に用い、同じ $\hat{\mathscr{P}}$ を評価に用いれば得られる。
この手続きを再代入法 (Resubstitution/R 法) と呼ぶ。

再代入法 (Resubstitution/R 法)

1. 与えられたデータセット S を用いて識別器を学習する
2. 同じセットを用いて性能を評価する

交差確認法 (Cross-Validation/CV 法)

H 法はデータを学習用と評価用に分割する必要があるため、誤り推定に不利な点がある。
推定の偏りは学習データ量の影響を受け、推定の分散は評価データ量の影響を受ける。
この問題点を解決する方法として、交差確認法 (Cross-Validation/CV 法) が利用される。

交差確認法 (Cross-Validation/CV 法)

1. 与えられたデータセット S を N の互いに素の集合 $S_1, S_2, \cdots, S_N$ に分割する
2. 一つの集合を評価セットとし、残りの $N-1$ をあわせて学習セットとし、学習と誤り推定を行う
3. 評価セットを入れ替えながら上記を N 回繰り返し、得られる N 個の誤り率を平均する

交差確認法 (Cross-Validation/CV 法)

CV 法では、データセットの各要素を個別の集合として扱うこともできる。
この極端な事例は一つ抜き法 (leave-one-out/L 法) と呼ばれる。

ブートストラップ法

ベイズ誤り率 ε の推定において、まず R 法にて推定値 $\hat{\varepsilon}$ を得ることができる。
次いでその偏りを考えよう

$$b = \varepsilon - \hat{\varepsilon}.$$

もし b の適切な推定値を得ることができれば、以下によりベイズ誤り率を推定することができる

$$\varepsilon = \hat{\varepsilon} + b.$$

ブートストラップ法

ブートストラップ法 (bootstrap method) は、元の集合 S から、復元抽出により仮想集合 S^* を得る ($|S| = |S^*|$ とする)。
次いで偏り b を以下により推定する

$$b^* = \varepsilon^* - \hat{\varepsilon}^*$$

ただし ε^* は S^* を学習セット S を評価セットとして得た誤りの推定値であり、$\hat{\varepsilon}^*$ は S^* を学習セット並びに評価セットとして得た誤りの推定値である。
これを B 回、B 個の仮想集合を生成して繰り返し、その平均値を得ることで b^* を推定する。
ベイズ誤りは以下のように推定する

$$\varepsilon = \hat{\varepsilon} + b^*$$

ブートストラップ法

ブートストラップ法 (The Bootstrap Method)

1. 与えられたデータセット S を学習用と評価用に用い (R 法) 識別誤り率の推定値 $\hat{\varepsilon}$ を得る
2. S からの $|S|$ 回の復元抽出により仮想セット S^* を得る
3. S^* を学習セット、S を評価セットとして、識別誤り率の推定値 ε^* を得る
4. S^* を学習セット並びに評価セットとして用いて識別誤り率の推定値 $\hat{\varepsilon}^*$ を得る
5. 偏りの推定値を算出する: $b^* = \varepsilon^* - \hat{\varepsilon}^*$
6. 仮想セットを都度生成しながらこの処理を B 回繰り返し、得られた B 個の偏りを平均して最終的な偏り b^* を得る
7. 誤り率の推定値を $\varepsilon = \hat{\varepsilon} + b^*$ で算出する

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

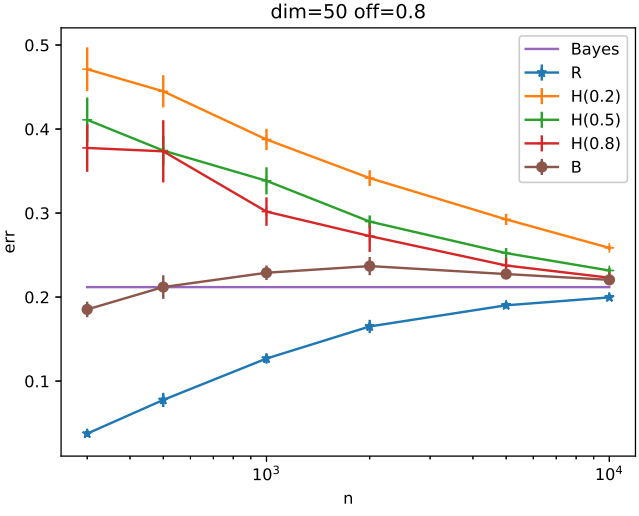
Class 2 $\Sigma_2 = I,\ \mu_2 = [-s, 0, \cdots, 0]^T$

ベイズ誤り率は

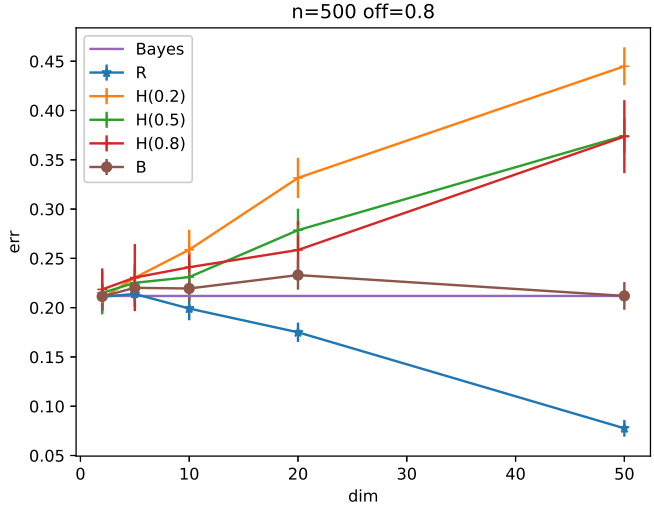
$$
\begin{aligned}
P^* &= \int_s^\infty \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty N(0, I) dx_1 \cdots dx_n \\
&= \frac{1}{2}(1 - \mathrm{erf}(\frac{s}{\sqrt{2}})) \\
\mathrm{erf}(x) &= \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt
\end{aligned}
$$

クラス 1 と 2 からそれぞれ n サンプル得たとして、二次識別器の誤り率を R 法と H 法で算出し、ベイズ誤り率と比較してみよう

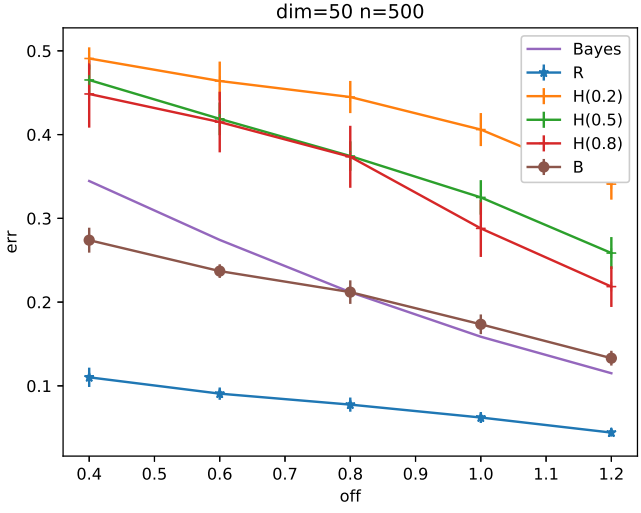
Exercise



Exercise


n=500 off=0.8
Bayes
R
H(0.2)
H(0.5)
H(0.8)
B
err
dim
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
0
10
20
30
40
50

Exercise


dim=50 n=500
Bayes
R
H(0.2)
H(0.5)
H(0.8)
B
err
0.5
0.4
0.3
0.2
0.1
0.4
0.5
0.6
0.7
0.8
0.9
1.0
1.1
1.2
off

宿題 (プログラミング課題)

- ブートストラップ法を実装せよ
- 演習の方法に従い、誤り率の推定結果を R 法と H 法並びにベイズ誤り率と比較せよ
- データの次元数、ベイズ誤り率、データセットの標本数を変えた時のふるまいを図示せよ
- 締切は 5 月 30 日
- PDF フォーマットにて ITC LMS に提出すること。
- Jupyter Notebook (ipynb) や Python (py) 形式の提出はおやめください。