ベイズ決定則・ロジスティック回帰・線形識別器

本谷 秀堅

名古屋工業大学

知的な振る舞いをする機械

外界を計測し、その状態を予測し，予測結果に応じて動作する

- 回帰：「大小関係に意味のある値」を予測
  - これまでの株価から明日の【株価】
  - 身長から【体重】
  - 顔画像から【年齢】
- 認識：「大小関係に意味の無い離散値」を予測
  - 顔画像から【本人か否か】【氏名】
  - 検索語から【関係の有無】
  - 音声から【音素】

本講義のテーマは【認識機械の作り方】

例題: ２クラス識別問題

髪の毛の長さから、その人物の性別を識別する機械を作りなさい：

- 入力：髪の毛の長さ, x。
- 出力：性別, $y \in \{1, 0\}$。 ($y = 1$(男性) or $y = 0$(女性))

定式化：2クラス識別機械の構築

目の前の人物が男性なら正に、女性なら負になるような 識別関数 $f(x)$ を作りなさい。

$$\hat{y}(x) = \begin{cases} 1, & \text{if } f(x) > 0, \\ 0, & \text{if } f(x) < 0. \end{cases}$$

識別の実際

1. x が入力される
2. $f(x)$ を計算して符号を調べる
3. カテゴリの推定値, $\hat{y}(x)$, を出力する

識別器を作る際の方針

識別境界は　　　　　が最小になるように決める。

誤識別率 (2 クラス識別問題)

データ $\boldsymbol{x}$ を誤識別する確率は次式のとおり：

$$P(\text{error}|\boldsymbol{x}) = \begin{cases} P(y=0|\boldsymbol{x}), & \text{if } \hat{y}(\boldsymbol{x}) = 1 \\ P(y=1|\boldsymbol{x}), & \text{if } \hat{y}(\boldsymbol{x}) = 0 \end{cases}$$

平均誤識別率

平均誤識別率は次式のとおり：

$$P(\mathsf{error}) = \int_{-\infty}^{\infty} \qquad\qquad d\boldsymbol{x}$$

識別機械は　　　　　　　　　が最小となるように設計する。平均誤識別率は次式のように計算できる

$$P(\mathsf{error}) = \int \qquad P(y=1|\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x} + \int \qquad P(y=0|\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}$$

ただし，$\boldsymbol{x} \in \Omega_0$ のとき $\hat{y}(\boldsymbol{x}) = 0$ であり，$\boldsymbol{x} \in \Omega_1$ のとき $\hat{y}(\boldsymbol{x}) = 1$ であるとする。

条件付きベイズ誤り確率

条件付きベイズ誤り確率

各 $\boldsymbol{x}$ に対する誤識別の理論的な最小値：

$$e_{\mathrm{B}}(\boldsymbol{x}) = \min\{P(y=0|\boldsymbol{x}), P(y=1|\boldsymbol{x})\}$$

$e_{\mathrm{B}}(\boldsymbol{x})$ は分布の重なりに由来する「必然的な誤り」

ベイズ誤り確率

ベイズ誤り確率

平均誤識別率の理論的な最小値

$$P_{\mathrm{B}}(\text{error}) = \int e_{\mathrm{B}}(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x} = \int \min\{P(C_1|\boldsymbol{x}), P(C_2|\boldsymbol{x})\}p(\boldsymbol{x})d\boldsymbol{x}$$

ベイズ誤り確率 (多クラスの場合)

K-クラス識別：$y \in \{0, 1, 2, \ldots, K-1\}$

条件付きベイズ誤り確率

$$e_{\mathrm{B}}(\boldsymbol{x}) = \min_{y}\{1 - P(y|\boldsymbol{x})\}$$

ベイズ誤り確率

$$P_{\mathrm{B}}(\textsf{error}) = \int \min_{y}\{1 - P(y|\boldsymbol{x})\} p(\boldsymbol{x}) d\boldsymbol{x}$$

認識機械の作り方

ベイズ決定則

与えられた $\boldsymbol{x}$ に対する条件付き誤識別率を最小にするクラス推定：

$$\hat{y}(\boldsymbol{x}) = \arg\max_{y}\{P(y|\boldsymbol{x})\}$$

例題

男性 ($y=1$) と女性 ($y=0$) の人口比は同じで、事前確率について $P(y=1)=P(y=0)=1/2$ が成立しているとする。また、髪の毛の長さ x の条件付き確率密度関数は次式で与えられるとする ($\theta > 0$)。

$$P(x|y=1) = \begin{cases} x/36, & 0 \le x \le 6 \\ -x/36 + 1/3, & 6 \le x \le 12 \\ 0, & \text{otherwise} \end{cases}$$

$$P(x|y=0) = \begin{cases} (x-\theta)/36, & \theta \le x \le \theta + 6 \\ -(x-\theta)/36 + 1/3, & \theta + 6 \le x \le \theta + 12 \\ 0, & \text{otherwise} \end{cases}$$

例題

1. 男性・女性の髪の毛の長さの期待値を求めよ

$$\mathbb{E}_{x \sim P(x|y=1)}[x] =$$

$$\mathbb{E}_{x \sim P(x|y=0)}[x] =$$

2. $\theta = 6$ とする。ある人物の髪の毛の長さが 8cm である確率を求めよ

3. $\theta = 6$ とする。ある人物の髪の毛の長さを測ったら 8cm であった。その人物が男性（女性）であった確率を求めよ。

例題

1. $p(x|y=1)$ と $p(x|y=0)$ は、先の例題と同じとする $(\theta = 6)$。下記それぞれの状況で最適な閾値を求めよ。
   - デパートの客は、$p(y=0) = p(y=1) = 1/2$ である。デパートの客の性別を髪の毛の長さで識別するときの識別関数を求めよ。
   - $p(C_1) = p(C_2) = 1/2$ のときの $p(C_1|x)$ のグラフを描け。（横軸 x, 縦軸 $p(C_1|x)$）
   - 名工大学生は、$p(y=1) = 9/10, p(y=0) = 1/10$ である。名工大学生の性別を髪の毛の長さで識別するときの識別関数を求めよ。
2. $K = 3$ であり、$P(y=0|\boldsymbol{x}_{\mathrm{obs}}) = 1/2$, $P(y=1|\boldsymbol{x}_{\mathrm{obs}}) = 1/3$, $P(y=2|\boldsymbol{x}_{\mathrm{obs}}) = 1/6$ であった。
   1. $\boldsymbol{x}_{\mathrm{obs}}$ に対する誤り確率が最小となるのは、$\boldsymbol{x}_{\mathrm{obs}}$ をどのクラスであると判定するときか述べよ。
   2. $e(\boldsymbol{x}_{\mathrm{obs}})$ を求めよ

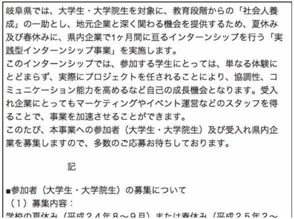
復習：最尤推定・生成モデル・ロジスティック回帰

メールが SPAM$(y=1)$ か否か $y=0$ を識別する 2 クラス識別機械を作れ。

- 入力：メール, $\boldsymbol{x}$
- 出力：SPAM か否か $(y \in \{0,1\})$

生成モデルによる識別器の構築

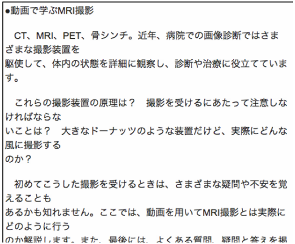
単語辞書を用意して、使われた単語に 1 、それ以外に 0 を付すことでメールを表現する。



岐阜県では、大学生・大学院生を対象に、教育段階からの「社会人養成」の一助とし、地元企業と深く関わる機会を提供するため、夏休み及び春休みに、県内企業で1ヶ月間に亘るインターンシップを行う「実践型インターンシップ事業」を実施します。
このインターンシップでは、参加する学生にとっては、単なる体験にとどまらず、実際にプロジェクトを任されることにより、協調性、コミュニケーション能力を高めるなど自己の成長機会となります。受入れ企業にとってもマーケティングやイベント運営などのスタッフを得ることで、事業を加速させることができます。
このたび、本事業への参加者（大学生・大学院生）及び受入れ県内企業を募集しますので、多数のご応募お待ちしております。

記

■参加者（大学生・大学院生）の募集について
（１）募集内容：
学校の夏休み（平成２４年８～９月）または春休み（平成２５年２～



●動画で学ぶMRI撮影

CT、MRI、PET、骨シンチ。近年、病院での画像診断ではさまざまな撮影装置を
駆使して、体内の状態を詳細に観察し、診断や治療に役立てています。

これらの撮影装置の原理は？ 撮影を受けるにあたって注意しなければならな
いことは？ 大きなドーナッツのような装置だけど、実際にどんな風に撮影する
のか？

初めてこうした撮影を受けるときは、さまざまな疑問や不安を覚えることも
あるかも知れません。ここでは、動画を用いてMRI撮影とは実際にどのように行う
のか解説します。また、最後には、よくある質問、疑問と答えを掲

x=

岐阜県	
動画	
プロジェクト	
ドーナツ	
企業	
大学生	
撮影	
装置	
MRI	

$\boldsymbol{x}$ の次元（単語数) は、10 万以上

準備：単純ベイズと条件付き確率

- $\boldsymbol{x}_i = [x_{i1}, x_{i2}, \ldots, x_{id}]^\top$
- SPAM メールが届く確率：$p(y_\cdot = 1) = \pi$
- 非 SPAM メールが届く確率：$p(y_\cdot = 0) = 1 - \pi$
- SPAM メールに単語 j が含まれる確率：$p(x_{\cdot j} = 1 | y_\cdot = 1) = p_j$
- 非 SPAM メールに単語 j が含まれる確率：$p(x_{\cdot j} = 1 | y_\cdot = 0) = q_j$

独立性の強い仮定 (　　　　　　　) の下での条件付き確率

$$p(\boldsymbol{x}_i | y = 1) = \prod_{j=1}^{d} p(x_{ij} | y = 1) =$$

$$p(\boldsymbol{x}_i | y = 0) = \prod_{j=1}^{d} p(x_{ij} | y = 0) =$$

復習：最尤法による $\pi = p(y)$ の最尤推定

学習データの集合, $\mathcal{D}$ を用意する

$$\mathcal{D} = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_n, y_n)\}$$

ベルヌーイ分布を採用し，$\pi = p(y = 1)$ を最尤推定する

$$p(y = 1) = \pi, \quad p(y = 0) =$$

$$p(y) =$$

π の最尤推定：

$$\hat{\pi}_{\mathrm{ML}} =$$

復習：p_j, q_j の最尤推定

$p_j = p(x_{\cdot j}|y_{\cdot} = 1)$ の最尤推定

$$\hat{p}_j =$$

$q_j = p(x_{\cdot j}|y_{\cdot} = 0)$ の最尤推定

$$\hat{p}_j =$$

復習：事後確率

届いたメール $\boldsymbol{x}$ が SPAM である事後確率:

$$p(y=1|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|y=1)p(y=1)}{p(\boldsymbol{x})}$$

$$=$$

復習：事後確率のシグモイド関数による表現

事後確率をロジスティックシグモイド関数で表す：

$$p(y = 1|\boldsymbol{x}) = \sigma(a),$$

ただし、$\sigma(a)$ はロジスティックシグモイド関数

$$\sigma(a) = \frac{1}{1 + \exp(-a)}.$$

logit 関数（ロジスティック関数の逆関数）

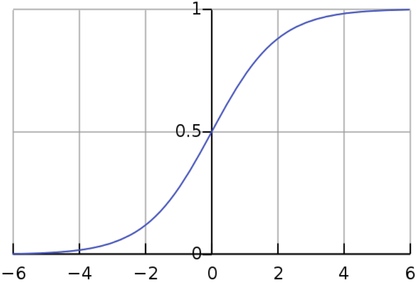
$$a =$$

整理すると次のとおり

$$a =$$

識別関数とロジット関数

線形識別関数が得られた。



$$f(\boldsymbol{x}) = a = w_0 + \sum_{j=1}^{d} w_j x_j = w_0 + \boldsymbol{w}^\top \boldsymbol{x}.$$

演習

学習データ $\{(y_i, \boldsymbol{x}_i) | i = 1, 2, \cdots, 8\}$ は下記の表のとおりであった。
$\boldsymbol{x}_i = [x_{i1}, x_{i2}, x_{i3}]^T$ であり、y, x_i はいずれも 0 or 1 の二値変数である。 ここまでの内容を利用して線形識別器を作れ。

	y	x_1	x_2	x_3
x^1	1	0	0	0
x^2	1	0	0	0
x^3	1	0	1	0
x^4	1	1	1	1
x^5	0	1	1	1
x^6	0	1	1	0
x^7	0	1	0	0
x^8	0	0	0	0

生成モデル

SPAM フィルタの作成：ここまでの粗筋

1. $\pi = P(y = 1)$, $p_j = P(x_{\cdot j}|y_{\cdot} = 1)$, $q_j = P(x_{\cdot j}|y_{\cdot} = 0)$ を最尤推定
2. ロジット関数を求めることで線形識別器が得られた
   - 線形識別関数の係数は π, p_j, q_j から求めた

メールデータを擬似的に生成できる確率モデルを獲得済み

$$p(\boldsymbol{x}) = p(y = 1)p(\boldsymbol{x}|y = 1) + p(y = 0)p(\boldsymbol{x}|y = 0)$$

推定したパラメータの数は　　　　個

ロジスティック回帰による識別器の構築

復習：ロジスティック回帰 (1/2)

与えられている学習データ：
$\boldsymbol{x}_{ij} \in \{0, 1\}$ $(j = 1, 2, \ldots, d)$,
$y_i \in \{0, 1\}$ $(i = 1, 2, \ldots, n)$

$$\mathcal{D} = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_n, y_n)\}$$

事後確率 $p(y = 1|\boldsymbol{x})$ をシグモイド関数で表す

$$r(\boldsymbol{w}) = p(y = 1|\boldsymbol{x}) = \sigma(a) = \frac{1}{1 + \exp(-a)}$$

事後確率 $p(y|\boldsymbol{x})$ の表現にはベルヌーイ分布を採用する

$$p(y|\boldsymbol{x}) = r^y (1 - r)^{1-y} = \left(\frac{1}{1 + \exp(-a)}\right)^y \left(1 - \frac{1}{1 + \exp(-a)}\right)^{1-y}$$

復習：ロジスティック回帰 (2/2)

ロジット関数を線形にする：$a = w_0 + \boldsymbol{w}^\top \boldsymbol{x}$

$$p(y|\boldsymbol{x}) = \left(\frac{1}{1+\exp\{-(w_0+\boldsymbol{w}^\top\boldsymbol{x})\}}\right)^y \left(\frac{\exp\{-(w_0+\boldsymbol{w}^\top\boldsymbol{x})\}}{1+\exp\{-(w_0+\boldsymbol{w}^\top\boldsymbol{x})\}}\right)^{1-y}$$

学習データ D に対する $\boldsymbol{w}$ の対数尤度関数

$$L(w_0, \boldsymbol{w}) = \log \prod_{i=1}^{n} p(y_i|\boldsymbol{x}_i)$$

$$\hat{w}_0, \hat{\boldsymbol{w}} = \arg\max_{w_0, \boldsymbol{w}} L(w_0, \boldsymbol{w})$$

ロジスティック回帰による識別器の構築

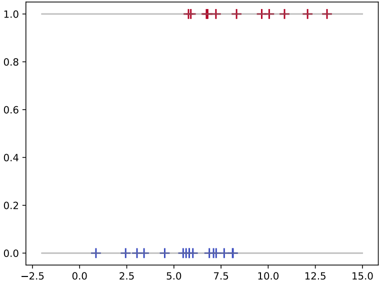
識別関数が得られた

$$f(\boldsymbol{x}) = \hat{w}_0 + \hat{\boldsymbol{w}}^\top \boldsymbol{x}$$

推定したパラメータ数は　　　個。
擬似データを生成可能にする情報は得られていない。

ロジスティック回帰による識別器の構築

例：1 次元

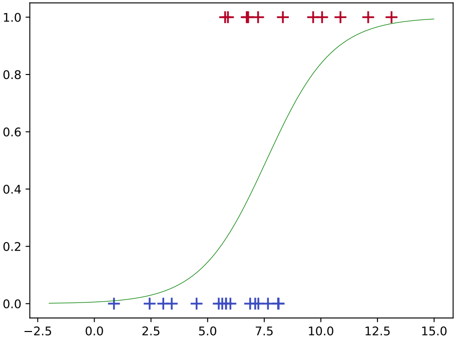


$$r(w_0, w_1) = p(y = 1|x) = \frac{1}{1 + \exp\{-(w_0 + w_1 x)\}}$$

ロジスティック回帰による識別器の構築

$$\hat{w}_0, \hat{w}_1 = \arg\max_{\boldsymbol{w}} L(w_0, w_1)$$

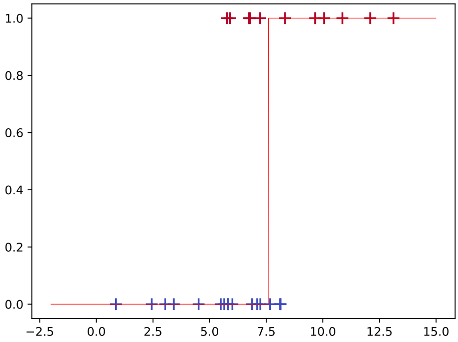
得られる曲線は各 x における $p(y = 1|x)$ の回帰結果



$$w_0 = -5.20, w_1 = 0.68$$

ロジスティック回帰による識別器の構築

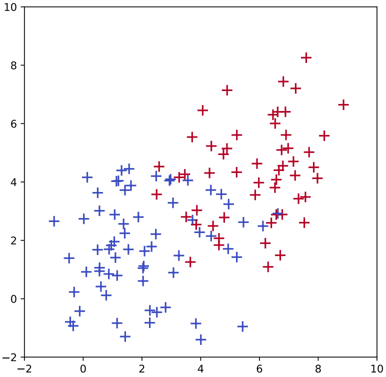
事後確率の大小で識別すると次のとおりの識別器が得られる



線形識別関数：$f(x) = -5.20 + 0.68x$

ロジスティック回帰による識別器の構築（多次元）

赤と青の点が分布している



新しい点が赤か青かを識別する機械を作れ

ロジスティック回帰による識別器の構築（多次元）

$\boldsymbol{w} = (w_0, w_1, w_2)^T$, $\boldsymbol{x} = (1, x_1, x_2)$,

$$r(\boldsymbol{w}) = p(\text{赤}\,|\boldsymbol{x}) = \frac{1}{1 + \exp(-\boldsymbol{x}^T\boldsymbol{w})}$$

赤を $y = 1$, 青を $y = 0$ としてベルヌーイ分布に基づいた対数尤度を式で表すと、前の例と同じになる

$$L(\boldsymbol{w}) = \log \prod_{i=1}^{N} p(y_i|\boldsymbol{x}_i)$$

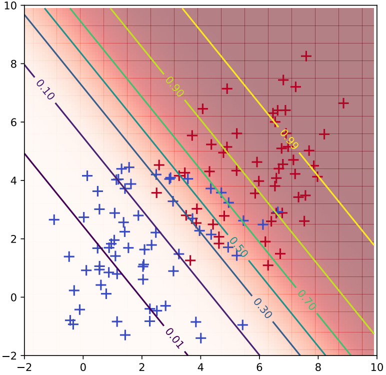
$$= \sum_{i=1}^{N} \left(y_i \log \frac{1}{1 + \exp\{-(\boldsymbol{x}_i^T\boldsymbol{w})\}} + (1 - y_i) \frac{\exp\{-(\boldsymbol{x}_i^T\boldsymbol{w})\}}{1 + \exp\{\exp\{-(\boldsymbol{x}_i^T\boldsymbol{w})\}} \right)$$

ロジスティック回帰は次式を計算する

$$\hat{\boldsymbol{w}} = \arg\max_{\boldsymbol{w}} L(\boldsymbol{w})$$

ロジスティック回帰による識別器の構築（多次元）

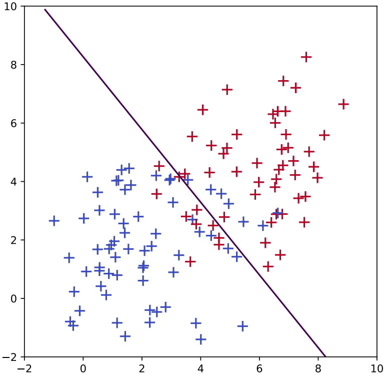
$$r(\hat{\boldsymbol{w}})$$



$$w_0 = -6.52,\ w_1 = 0.98,\ w_2 = 0.79$$

ロジスティック回帰による識別器の構築（多次元）

識別関数：$f(\boldsymbol{x}) = -6.52 + 0.98x_1 + 0.79x_2$



識別境界は $f(\boldsymbol{x}) = 0$

一般化線形モデル

一般化線形モデル, $r = g(\boldsymbol{w}^T \boldsymbol{x})$, による回帰

活性化関数とリンク関数

- 一般化線形モデルにおける $g(\cdot)$ を活性化関数と呼ぶ
- 活性化関数の逆関数 $g^{-1}(\cdot)$ をリンク関数 (連結関数) と呼ぶ

活性化関数は非線形なのに「線形モデル」と呼ぶ。なぜなら、
等高線, $r = g(\boldsymbol{w}^T \boldsymbol{x}) =$ 一定、が線形になるから。

一般化線形モデル

準備：$a = \boldsymbol{w}^\top \boldsymbol{x} = w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_d x_d$ である。w_0 とそれ以外を分けて次のように表す。

$$a = w_0 + \boldsymbol{\omega}^\top \boldsymbol{x}$$

ただし、$\boldsymbol{\omega} = (w_1, w_2, \ldots, w_d)^\top$

一般化線形モデル

$$a = w_0 + \boldsymbol{\omega}^\top \boldsymbol{x} = a_{\text{const}}$$

$$\left(\frac{\boldsymbol{\omega}}{\|\boldsymbol{\omega}\|}\right)^\top \boldsymbol{x} = \frac{a_{\text{const}} - w_0}{\|\boldsymbol{\omega}\|}$$

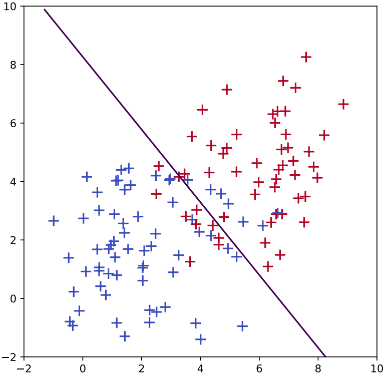
$$\|\boldsymbol{x}\| \cos\theta = \frac{a_{\text{const}} - w_0}{\|\boldsymbol{\omega}\|}$$

一般線形モデル

$f(a) = r_{\text{const}}$ の等高線は超平面になり、
$a_{\text{const}} = f^{-1}(r_{\text{const}})$ と表すとき、
その超平面の法ベクトルは $\boldsymbol{\omega}$ であり、
原点からの距離は $(a_{\text{const}} - w_0)/\|\boldsymbol{\omega}\|$ である。

実験結果の確認

識別関数：$f(\boldsymbol{x}) = -6.52 + 0.98x_1 + 0.79x_2$



識別境界は $f(\boldsymbol{x}) = 0$

ロジスティック関数の逆関数

ロジスティック関数は次式のとおり

$$r = \frac{1}{1 + \exp(-a)}$$

その逆関数は次式のとおり

$$a = \log \frac{r}{1 - r}$$

この関数をロジット関数 (logit function) という。

事後確率をロジスティック回帰 $r = p(y = 1|\boldsymbol{x})$ したとする。このとき次式が成り立つ

$$a = \log \left(\frac{p(y = 1|\boldsymbol{x})}{1 - p(y = 1|\boldsymbol{x})} \right) =$$

二つのクラスについての事後確率の比は「　　　」と呼ばれ、その対数は「　　　　　」と呼ばれる。識別関数にそのまま使える。