アジェンダ

1. 講義

a.アウトライン

b.数学的定式化

i.データ形式の説明

ii.ロジスティック回帰モデルの説明

iii.パラメータの推定

iv.モデルの評価

2.ハンズオン

a.タイタニックデータ解析

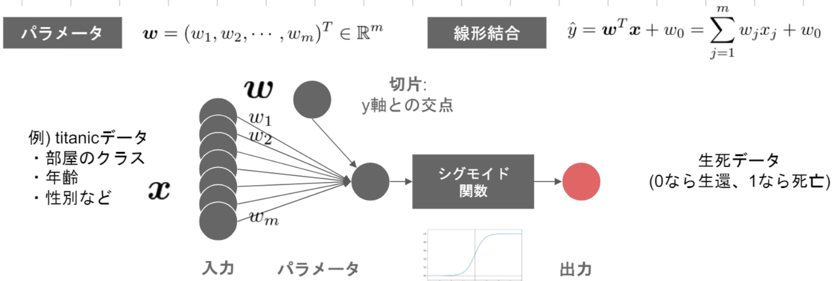
b.特徴量抽出分類タスクロジスティック回帰モデル

ロジスティック回帰モデル

分類問題（クラス分類）:入力~クラスに分類する問題。

入力：説明変数：x - (x1,x2・・・xm)T ：m次元のベクトル(m=1の場合はスカラ)

出力：目的変数：y∈{0 or 1}

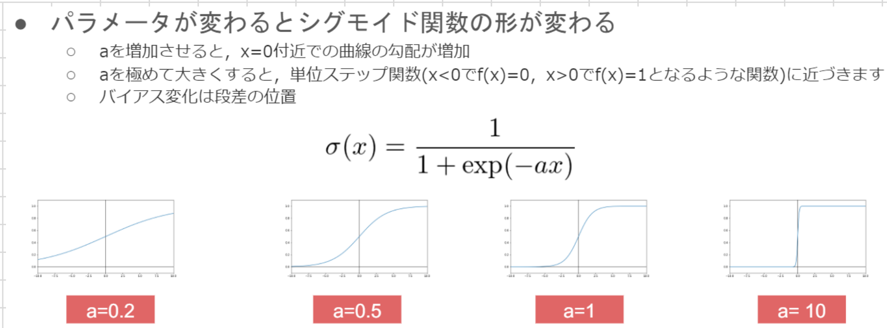


シグモイド関数

入力は実数・出力は必ず0~1の値

クラス1に分類される確率を表現

単調増加関数



シグモイド関数の微分は、シグモイド関数自身で表現することが可能。

尤度関数の微分を行う際にこの事実を利用すると計算が容易。

$$\frac{\partial \sigma(x)}{\partial x} = \frac{\partial}{\partial x}\left(\frac{1}{1+\exp(-ax)}\right)$$

$$= (-1) \cdot \{1+\exp(-ax)\}^{-2} \cdot \exp(-ax) \cdot (-a)$$

連鎖律

$$= \frac{a\exp(-ax)}{\{1+\exp(-ax)\}^2} = \frac{a}{1+\exp(-ax)} \cdot \frac{1+\exp(-ax)-1}{1+\exp(-ax)}$$

$$= a\sigma(x)(1-\sigma(x))$$

$$y = \frac{f(x)}{g(x)}$$ **の導関数は**

$$y' = \frac{f'(x)g(x) - f(x)g'(x)}{\{g(x)\}^2}$$

【指数関数の微分 公式1】

$$y = e^x \rightarrow y' = e^x$$

$$y = e^{kx} \rightarrow y' = ke^{kx}$$

sigmoid(x) = 1 / {1 + exp(-ax) }

sigmoid(x)' = a * sigmoid(x) * { 1 - sigmoid(x) }

ロジスティック回帰でのモチベーション：シグモイド関数を利用して、Y=1となるような関数を求めたい。

求めたい値　シグモイド関数

$$\underbrace{P(Y=1|\boldsymbol{x})}_{\text{説明変数の実現値が与えられた際にY=1になる確率}} = \sigma(\underbrace{w_0 + w_1x_1 + \cdots + w_mx_m}_{\text{データのパラメーターに対する線形結合}})$$

最尤推定

ベルヌーイ分布：コイン投げイメージの確率分布

最尤推定：得られたデータからもっともらしい母数分布を推定する方法→尤度関数を最大化する。

同時確立：あるデータが得られたときに、それが同時に得られる確率

尤度関数：データは固定し、パラメータを変化させる。

既知　未知

1回の試行で y=y_1になる確率　$P(y) = p^y(1-p)^{1-y}$

n回の試行で y1~ynが同時に起こる確率(p固定)　$P(y_1, y_2, \cdots, y_n; p) = \prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i}$

y1~ynのデータが得られた際の尤度関数　$P(y_1, y_2, \cdots, y_n; p) = \prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i}$

※尤もらしいPを求める！

確率pはシグモイド関数となるため、推定するパラメータは重みパラメータとなる

ロジスティック回帰モデルの最尤推定
- 確率pはシグモイド関数となるため、推定するパラメータは重みパラメータとなる
- $(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_1), \cdots, (\boldsymbol{x}_n, y_n)$ を生成するに至った尤もらしいパラメータを探す

★xiが具体的にイメージがわかない。
推察：
0 ＜ x ＜ 1 な固定値で(なんでも)よいのか？
0 ＜ x ＜ 1 な固定値をwiの学習で制御する考えか？

$$P(Y=y_1|\boldsymbol{x}_1) = p_1^{y_1}(1-p_1)^{1-y_1} = \sigma(\boldsymbol{w}^T\boldsymbol{x}_1)^{y_1}(1-\sigma(\boldsymbol{w}^T\boldsymbol{x}_1))^{1-y_1}$$
$$P(Y=y_2|\boldsymbol{x}_2) = p_2^{y_2}(1-p_2)^{1-y_2} = \sigma(\boldsymbol{w}^T\boldsymbol{x}_2)^{y_2}(1-\sigma(\boldsymbol{w}^T\boldsymbol{x}_2))^{1-y_2}$$
$$\cdots$$
$$P(Y=y_n|\boldsymbol{x}_n) = p_n^{y_n}(1-p_n)^{1-y_n} = \sigma(\boldsymbol{w}^T\boldsymbol{x}_n)^{y_n}(1-\sigma(\boldsymbol{w}^T\boldsymbol{x}_n))^{1-y_n}$$

：既知　：未知

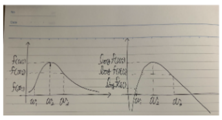
★尤度関数　**★推察：Pの引数は yiだけでなく、xiも含む。でよいか？（xiを与えるとyiとなるような wi を求めていく?）**

$$\begin{aligned}
P(y_1, y_2, \cdots, y_n | w_0, w_1, \cdots, w_m) &= \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i} \\
&= \prod_{i=1}^{n} \sigma(\boldsymbol{w}^T\boldsymbol{x}_i)^{y_i}(1-\sigma(\boldsymbol{w}^T\boldsymbol{x}_i))^{1-y_i} \\
&= L(\boldsymbol{w})
\end{aligned}$$

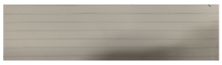
尤度関数はパラメータのみに依存する関数

★対数尤度関数

- 尤度関数を最大とするパラメータを探す (推定)
  - 対数をとると微分の計算が簡単
    - 同時確率の積が和に変換可能
    - 指数が積の演算に変換可能
  - 対数尤度関数が最大になる点と尤度関数が最大になる点は同じ
    - 対数関数は単調増加 (ある尤度の値がx1 < x2の時、必ずlog(x1) < log(x2)となる)
  - 「尤度関数にマイナスをかけたものを最小化」し、「最小2乗法の最小化」と合わせる



$$E(w_0, w_1, \cdots, w_m) = -\log L(w_0, w_1, \cdots, w_m)$$
$$= -\sum^{n} \{y_i \log p_i + (1-y_i)\log(1-p_i)\}$$



$$= -\sum_{i=1} \{y_i \log p_i + (1 - y_i) \log(1 - p_i)\}$$

対数公式: **log(AxB) = logA + logB**
log(A^B) = BlogA

■勾配降下法

対数尤度関数をパラメータで微分して0になる値を求める必要があるのだが、解析的にこの値を求めることは困難

$$\boldsymbol{w}(k+1) = \boldsymbol{w}^k - \eta \frac{\partial E(\boldsymbol{w})}{\partial \boldsymbol{w}}$$

★対数尤度関数の微分

対数尤度関数を係数とバイアスに関して微分

$$\frac{\partial E(\boldsymbol{w})}{\partial \boldsymbol{w}} = \sum_{i=1}^{n} \frac{\partial E_i(\boldsymbol{w})}{\partial p_i} \frac{\partial p_i}{\partial \boldsymbol{w}}$$

連鎖律

$$= \sum_{i=1}^{n} \left(\frac{y_i}{p_i} - \frac{1 - y_i}{1 - p_i} \right) \frac{\partial p_i}{\partial \boldsymbol{w}}$$

対数尤度関数の
pに関する微分

$$= \sum_{i=1}^{n} \left(\frac{y_i}{p_i} - \frac{1 - y_i}{1 - p_i} \right) p_i(1 - p_i)\boldsymbol{x}_i$$

シグモイド関数の微分

$$= -\sum_{i=1}^{n} (y_i(1 - p_i) - p_i(1 - y_i))\boldsymbol{x}_i$$

式を整理

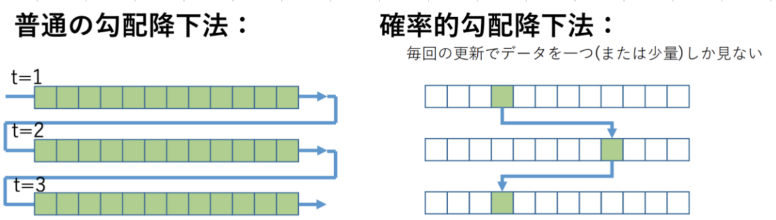
$$= -\sum_{i=1}^{n} (y_i - p_i)\boldsymbol{x}_i$$

確率的勾配降下法(SGD)

データを一つずつランダムに(「確率的」に)選んでパラメータを更新

勾配降下法でパラメータを1回更新するのと同じ計算量でパラメータをn回更新できるので効率よく最適な解を探索可能

$$\boldsymbol{w}(k+1) = \boldsymbol{w}^k + \eta(y_i - p_i)\boldsymbol{x}_i$$



評価

学習済みの「ロジスティック回帰モデル」の性能を測る指標 → 混同行列

正解率(Accuracy) **(TP + TN) / (TP + TN + FP + FN)**

再現率(Recall) **TP / (TP + FN)**

		ラベル (答え)	
		positive	negative
モデルの予測結果	positive	True Positive	False Positive
	negative	False Negative	True Negative

False Negative が小さい

適合率(Precision) **TP / (TP + FP)**

		ラベル (答え)	
		positive	negative
モデルの予測結果	positive	True Positive	False Positive
	negative	False Negative	True Negative

False Positive が小さい

★要するに精度を上げたい分布を母数にする

F値(F-measure):調和平均 （※逆数の平均の逆数）

	メリット	デメリット
正解率	最もシンプルで分かりやすい	クラスごとの評価データ数が著しく異なると不適切
再現率	取りこぼしを発見できる	過検知を発見できない
適合率	過検知を発見できる	取りこぼしを発見できない
F値	取りこぼし、過検知をどちらも均等に判断できる	数値の解釈が難しくなる

←過検知してよいケース
ウィルス感染検査など
過検知したら再検査すればいい。
←取りこぼしてよいケース
スパムメール検知など。
取りこぼしたら自分で分類する。