ロジスティック回帰

正田 備也

masada@rikkyo.ac.jp

予測問題の2種類（復習）

- 回帰(regression)＝数値の予測
- 分類(classification)＝クラスの予測
  - 今回から、こちらに取り組む

クラスの予測＝分類(classification)

- 分類 (classification)
  - 未知のinstanceを、複数のクラスのいずれかへグループ分け
  - その際、グループ分けがすでに済んでいるデータを利用する
    - グループ分けが済んでいる＝正解が分かっている

  - 例：スパムフィルタ、手書き数字認識、など

ロジスティック回帰による2値分類

- ロジスティック回帰は、名前が示す通り、回帰
- つまり、クラスの予測ではなく、<u>値の予測</u>
- だが、ロジスティック回帰が予測するのは、値は値でも、<u>確率</u>
- 確率とは、０から１の範囲におさまる値
- 確率を、2つのクラスのうちの一方に属する確率として使う
- そうすると、ロジスティック回帰は2値分類に使える
  - （他方のクラスに属する確率）$= 1 -$（一方のクラスに属する確率）

ロジスティック回帰の目的変数

- 線形回帰の目的変数は$-\infty$から$+\infty$までの値をとる
- ロジスティック回帰の目的変数y_iは$\mathbf{0}$か$\mathbf{1}$の値をとる
  - 一方のクラスに属するか属さないかの、2値なので。
- この$\mathbf{0}$か$\mathbf{1}$の値を、モデルが出力する$\mathbf{0}$から$\mathbf{1}$までの連続的な値（一方のクラスへの所属確率p_i）で予測する

ロジスティック回帰と線形回帰の違い

- モデルを表現する式の左辺が違う

$$\log\left(\frac{p_i}{1-p_i}\right) = a_0 + a_1 x_{i,1} + \cdots + a_d x_{i,d}$$

  - 線形回帰では、モデルの出力$f(\boldsymbol{x}_i)$がそのままターゲットy_iの予測値
- p_iはi番目のデータ点$\boldsymbol{x}_i$が一方のクラスに属する確率
- 左辺の関数はlogitやlog-oddsと呼ばれる関数
  - それぞれのクラスに属する確率の比の対数。
  - このlogit関数のように、目的変数とモデルの出力とのつながりを表す関数を「リンク関数」と呼ぶ。

モデルを表す式の変形

$$\log\left(\frac{p_i}{1-p_i}\right) = f(\boldsymbol{x}_i)$$

$$\log\left(\frac{p_i}{1-p_i}\right) = a_0 + a_1 x_{i,1} + \cdots + a_d x_{i,d}$$

$$\Updownarrow$$

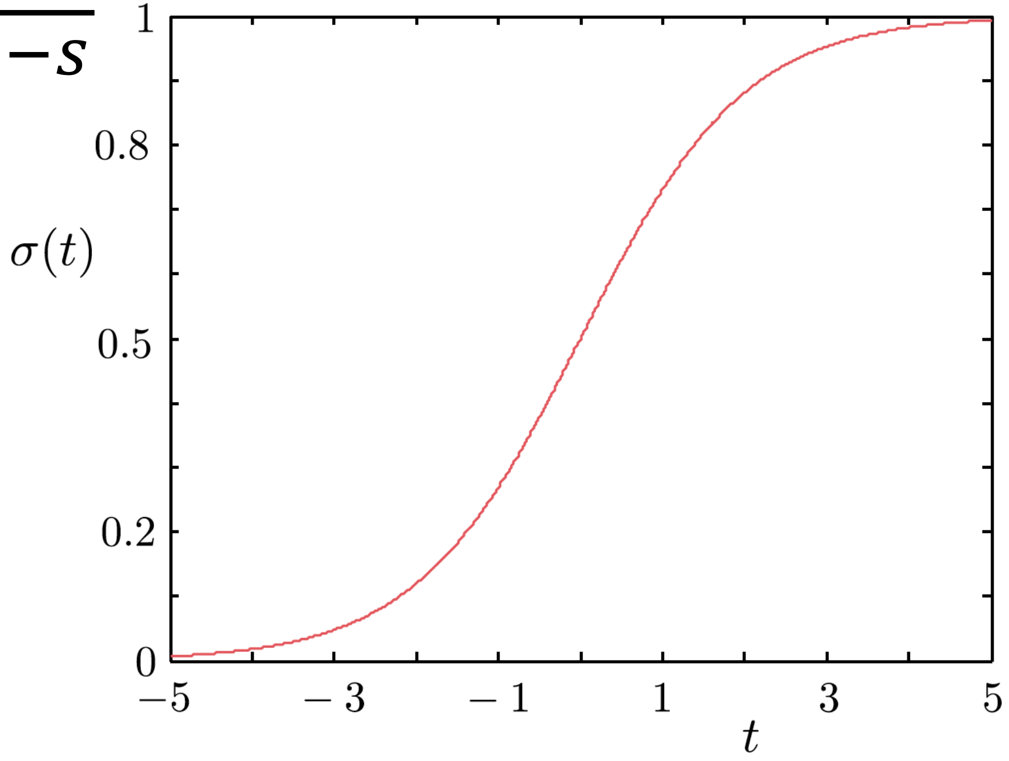
$$p_i = \frac{1}{1 + e^{-(a_0 + a_1 x_{i,1} + \cdots + a_d x_{i,d})}}$$

$$p_i = \frac{1}{1 + e^{-f(\boldsymbol{x}_i)}}$$

シグモイド関数

- logit関数の逆関数は、シグモイド関数と呼ばれる

$$\sigma(s) = \frac{1}{1+e^{-s}}$$



リンク関数の逆関数

- $\sigma(x)$をシグモイド関数とする
- ロジスティック回帰のモデリングをあらわす式より

$$p_i = \sigma(f(\boldsymbol{x}_i))$$

- $\boldsymbol{\sigma(x)}$のような関数をmean functionとも呼ぶ
  - 線形モデルの出力を変換し、目的変数につなげる関数
    - これが特定の確率分布の期待値に相当するため、こう呼ばれる。

ロジスティック回帰におけるモデリング

- モデルの出力$f(\boldsymbol{x}_i)$をシグモイド関数で確率p_iに変換
- このp_iが表の出る（1の出る）確率になっているコインを考える
- 目的変数の値y_iは、このコインを投げて生成されると考える
- 値 $y_1, \ldots, y_N$ のすべてが、別の確率に基づくコイン投げではなく、このコイン投げによって生成されると考えると一番尤もらしくなるように、パラメータ$a_0, a_1, \ldots, a_d$の値を推定する！
    - 各$\boldsymbol{x}_i$についてp_iは異なる値であることに注意。

線形回帰におけるモデリング

- モデルの出力$f(\boldsymbol{x}_i)$をそのまま平均とする正規分布を考える
  - ロジスティック回帰の場合のような変換は、しない。
- 目的変数の値y_iは、この正規分布から生成されると考える
- 値$y_1, \dots, y_N$のすべてが、別の正規分布ではなく、この正規分布から生成されると考えると一番尤もらしくなるように、パラメータ$a_0, a_1, \dots, a_d$の値を推定する
  - ただし、これは最小二乗法の説明。

一般化線形モデル

- 一般化線形モデルの部品
  1. 線形予測子
     - $f(\boldsymbol{x}_i) = a_0 + a_1 x_{i,1} + \cdots + a_d x_{i,d}$のこと。
  2. リンク関数
     - 線形回帰では恒等関数、ロジスティック回帰ではlogit。
  3. 確率分布
     - 線形回帰では正規分布、ロジスティック回帰ではベルヌーイ分布。

https://eprints.lib.hokudai.ac.jp/dspace/bitstream/2115/49477/4/kubostat2008c.pdf

ロジスティック回帰の損失関数

- 正解y_iは1か0の値しか取らない
- ロジスティック回帰の損失関数は以下のように書ける

$$L(a_0, a_1, \dots, a_d) = -\sum_{i=1}^{N}\{y_i \log p_i + (1 - y_i)\log(1 - p_i)\}$$

これは、何？

ただし

$$p_i = \sigma\left(a_0 + \sum_{j=1}^{d} a_j x_{i,j}\right) = \frac{1}{1 + e^{-(a_0 + a_1 x_{i,1} + \cdots + a_d x_{i,d})}}$$

クロスエントロピー

- 線形回帰の場合、モデルの出力と目標値とのズレを、差の二乗の和（ないし平均）で測っていた
- ロジスティック回帰の場合は、また別のズレの測り方をする
- それがクロスエントロピー
- i番目の個体についてモデルが予測した確率をp_iとし、ターゲットとなる1か0の値をy_iとすると

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

2値分類問題をロジスティック回帰で解く

- 2値分類問題の場合、正解は1か0になっている
  - あるクラスに属するか属さないかのいずれかで、中間は無いから。
- よって、クロスエントロピーで表された正解からのズレは
  - 正解y_iが1のときは

$$-\log p_i$$

  - 正解y_iが0のときは

$$-\log(1-p_i)$$

クロスエントロピーの正体

- 尤度の対数をとってマイナスをつけたもの
- この場合の尤度＝確率p_iで表が出る（1が出る）コインを投げたときにy_iが出る確率
  - 正解y_iが1のとき、y_iが出る確率はp_i
  - 正解y_iが0のとき、y_iが出る確率は$1 - p_i$
  - 両方のケースをまとめて書くと${p_i}^{y_i}(1 - p_i)^{(1 - y_i)}$
  - 対数をとると$y_i \log p_i + (1 - y_i) \log(1 - p_i)$
- クロスエントロピーの最小化＝対数尤度の最大化

最尤推定

- 確率モデルのパラメータの値を、データの尤度が最大値となるような値として推定する、パラメータ推定手法

- ロジスティック回帰の場合
  - 各入力ベクトルx_iに対応する正解ラベルy_iの尤度${p_i}^{y_i}(1-p_i)^{(1-y_i)}$を、すべての入力ベクトルについて掛け算してまとめる。

$$\prod_{i=1}^{N} {p_i}^{y_i}(1-p_i)^{(1-y_i)}$$

  - この掛け算した値が最大値をとるように、 $a_0, a_1, \ldots, a_d$を推定する。

損失関数の勾配の計算

- ライブラリにまかせましょう
  - 数学が得意な方は手計算を試みましょう。
- 知っておいた方がいいこと：シグモイド関数の微分

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

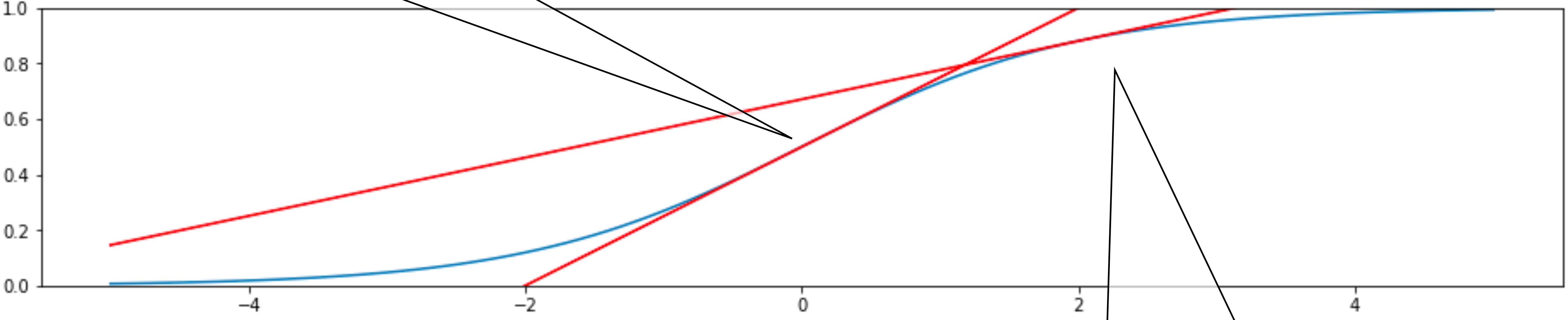
$$\frac{d}{dx}\sigma(x) = -\frac{-e^{-x}}{(1+e^{-x})^2} = \frac{e^{-x}}{(1+e^{-x})^2} = \frac{1}{1+e^{-x}}\left(1 - \frac{1}{1+e^{-x}}\right)$$

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

関数の値そのものを使って簡単に書ける！

$x = 0$での接線の傾きは

$$\sigma(0)\big(1 - \sigma(0)\big)$$
$$= 0.5(1 - 0.5) = 0.25$$



$x = 2$での接線の傾き

$$\sigma(2)\big(1 - \sigma(2)\big)$$
$$= 0.881(1 - 0.881)$$
$$= 0.105$$