ロジスティック回帰

機械学習演習 プランナークラス

masada@rikkyo.ac.jp

予測問題の二種類

- 回帰(regression)＝数値の予測
- 分類(classification)＝クラスの予測
  - 今回からこちら

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回帰＝<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
  - この確率を、2つのクラスのうちの一方に属する確率として使う
- だから、ロジスティック回帰は2値分類に使える
  - （もう一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におけるクラス予測

- ターゲットの値の範囲
  - 線形回帰：$-\infty$から$+\infty$まで
  - ロジスティック回帰：$\underline{1か0}$
    - 一方のクラスに属するか属さないかの2値

- ロジスティック回帰における予測
  - この1か0かの2値を「コイン投げ」で予測する

「コイン投げ」による予測

- 確率p_iで表が出るコインを投げてターゲットy_iを作る
  - ターゲットの値1がコインの表に、値0がコインの裏に、それぞれ対応

- 確率p_iは以下のように計算する

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

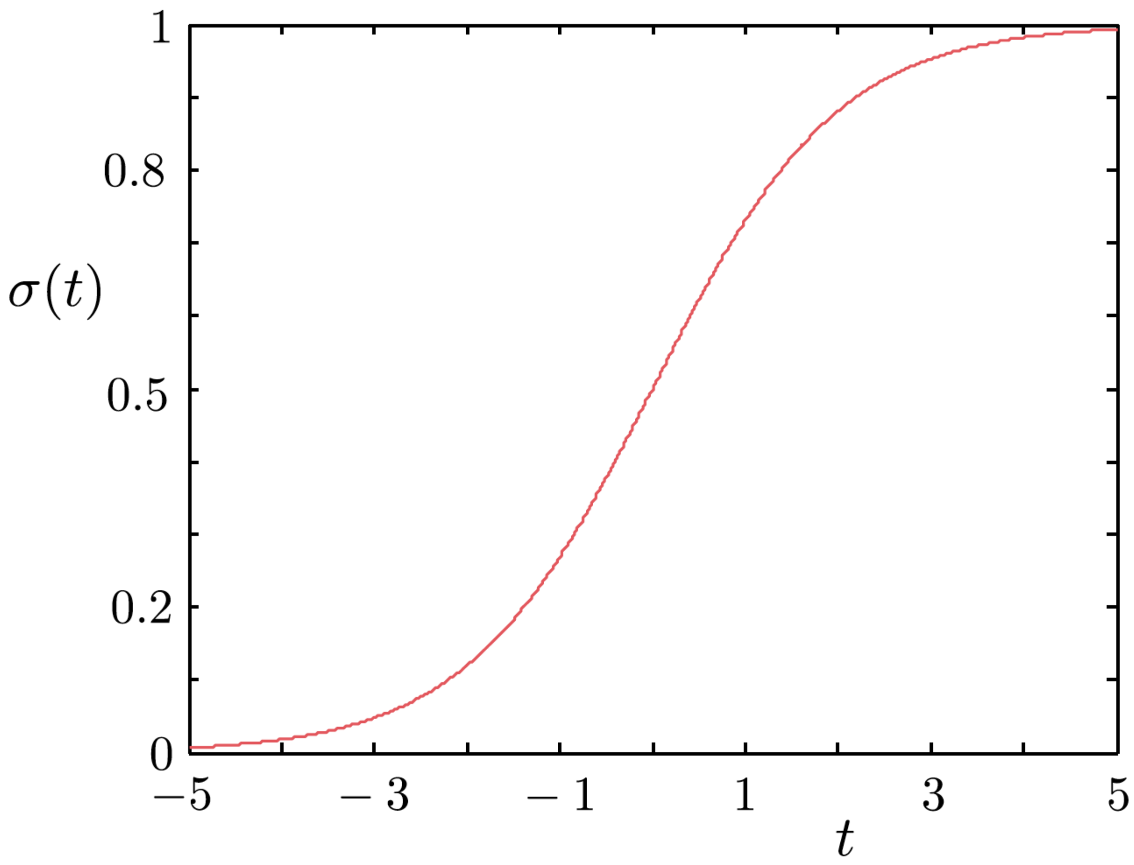
シグモイド関数

- シグモイド関数の式

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

- 確率p_iは以下のように書ける

$$p_i = \sigma(a_0 + a_1 x_{i,1} + \cdots + a_d x_{i,d})$$



ロジスティック回帰の予測モデル

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

$$\Updownarrow$$

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

ロジスティック回帰と線形回帰の違い

- 線形回帰と比べると予測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$a_0 + a_1 x_{i,1} + \cdots + a_d x_{i,d}$がそのまま予測値になる

- 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。

ロジスティック回帰におけるパラメータ推定

- 確率p_iで表が出るコインを投げてターゲットy_iを作る
  - ターゲットの値1がコインの表に、値0がコインの裏に、それぞれ対応

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

- パラメータ推定の方法
  - 訓練データのターゲットy_iの全て$(i = 1, \ldots, N)$が、このコイン投げによって作られる確率を、できるだけ大きくするように、パラメータ$a_0, a_1, \ldots, a_d$の値を求める。

線形回帰におけるパラメータ推定

- ターゲットy_iは、$a_0 + a_1 x_{i,1} + \cdots + a_d x_{i,d}$を平均とする正規分布に従うと考える
  - ロジスティック回帰と違って$a_0 + a_1 x_{i,1} + \cdots + a_d x_{i,d}$をそのまま使う

- パラメータ推定の方法
  - 訓練データのターゲットy_iの全て$(i = 1, \ldots, N)$が、この正規分布から作られる確率を、できるだけ大きくするように、パラメータ$a_0, a_1, \ldots, a_d$の値を求める。

ターゲットが作られる確率

- 場合分けして考えると・・・
  - ターゲットが$y_i = 1$のとき、このターゲットが作られる確率はp_i
  - ターゲットが$y_i = 0$のとき、このターゲットが作られる確率は$1 - p_i$

- 両方の場合をまとめて書くと・・・

$$p_i^{y_i}(1 - p_i)^{(1-y_i)}$$

- 全てのインスタンスの確率は・・・

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

ロジスティック回帰の損失関数

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

ただし

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

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
     - $a_0 + a_1 x_{i,1} + \cdots + a_d x_{i,d}$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
       - 参考資料: https://kuboweb.github.io/-kubo/stat/2018/Ees/c/HOkubostat2018c.pdf

損失関数の勾配の計算

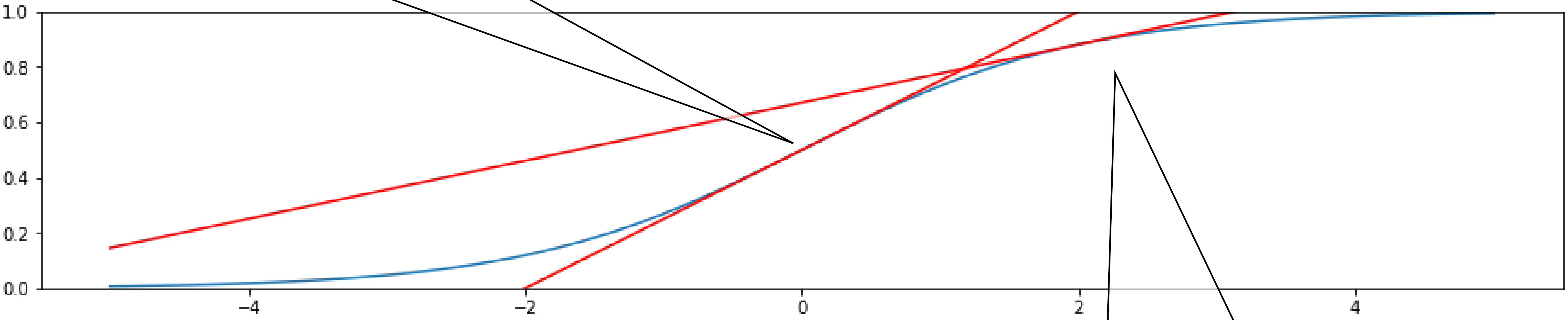
$$\sigma(x) = \frac{1}{1+e^{-x}}$$

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

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

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
$\sigma(0)\left(1 - \sigma(0)\right)$
$= 0.5(1 - 0.5) = 0.25$



$x = 2$での接線の傾き
$\sigma(2)\left(1 - \sigma(2)\right)$
$= 0.881(1 - 0.881)$
$= 0.105$