Logistic Regression (number between 0 & 1)
Saturday, 8 March 2025

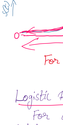
Propose: Estimate the probability that a point belongs to one of two classes $y \in \{a, b\}$

$x \in \mathbb{R}^d$
$b \quad y$

Logistic function (Sigmoid)

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

real number $\quad [0,1]$ $\quad$ exponential pun dim



$t \to 0$, $s(t) = \frac{e^0}{1+e^0} = \frac{1}{2}$

For $t >> 0$, $s(t) \approx 1$
$t << 0$, $s(t) \approx 0$

Logistic Regression

For a point a_i^T and its label $y \in \{0, 1\}$

$P(y=1|a_i^T) = \frac{1}{1+e^{-a_i^T x}} = s(a_i^T x)$ $\quad P(y=0|a_i^T) = 1 - P(y=1|a_i^T)$

$= \frac{e^{a_i^T x}}{1+e^{a_i^T x}} > 0.5$ $\quad t \to (a_i^T x)$

The output $s(a_i^T x)$ is interpreted as a probability

$\boxed{a_i^T x >> 0} \Rightarrow s(a_i^T x) >> 0.5$

$\Rightarrow P(y=1|a_i^T) \approx 1$

$\therefore a_i^T$ is classified as class 1

$a_i^T x << 0 \Rightarrow s(a_i^T x) << 0.5$

$\Rightarrow P(y=1|a_i^T) \approx 0$

$\therefore a_i^T$ is classified as 0.

Cost function / Loss

$\{0, 1\}$

For data $\{(a_1^T, y_1), (a_2^T, y_2), \ldots, (a_n^T, y_n)\}$

estimate

$P(y_i = 1 | a_i^T) = \frac{1}{1+e^{-a_i^T x}} = \pi_i$

π_i: prob th point belongs to label 1

$loss(x) = -\sum_{i=1}^{n} \left[\underbrace{y_i \times \ln \pi_i}_{y_i = 1} + \underbrace{(1-y_i) \ln(1-\pi_i)}_{y_i = 0} \right]$

Cost Function Interpretation

Suppose there is only one datapoint say (a_i^T, y_i)

$\therefore loss(x) = \begin{cases} -\ln \pi_i & \text{if } y_i = 1 \\ -\ln(1-\pi_i) & \text{if } y_i = 0 \end{cases}$

$y_i = 0$: $\ln(1-\pi_i)$, $-\ln(1-\pi_i)$

$y_i = 1$: $\ln \pi_i$, $-\ln \pi_i$

log of values between 0 & 1 is negative

$\ln \pi_i$ $-ve$ $-\ln \pi_i = +ve$

$\ln(1-\pi_i)$ $-ve$ $-\ln(1-\pi_i) = +ve$

Case $y_i = 1$

$loss(x) = -\ln \pi_i$ $\quad y_i = 1, 0$

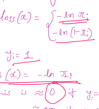
this is ≈ 0 if $y_i = 1$ and $\pi_i \approx 1$
$\approx +\infty$ if $y_i = 1$ and $\pi_i \approx 0$

0.2 $\quad \pi_i \leq 0.5$
$\pi_i = 1$
$P($

Case $y_i = 0$

$loss(x) = -\ln(1-\pi_i)$ $\quad \pi_i$

≈ 0 if $y_i = 0$ & $\pi_i = 0$
$\approx +\infty$ if $y_i = 0$ & $\pi_i = 1$



$\pi_i = p_i$

So, training means give data Matrix A & label vector y find $x \in \mathbb{R}^d$ s.t. $loss(x)$ is minimized

A, y

$$x^* = \underset{x \in \mathbb{R}^d}{\operatorname{argmin}} \; loss(x)$$

x^* is our model

For a new point a_{n+1}^T

let $P = \frac{1}{1+e^{-a_{n+1}^T x^*}}$

if $P \geq 0.5$, class $= 1$
else class $= 0$

→ One vs all for Multiclass classification

→ Since prediction still depends on $a^T x$ it is a generalized linear model

$\|Ax - b\|_2^2$ $\quad a_i^T x$ $\quad x_1 a_{i1} + x_2 a_{i2}$

$A \quad b \quad \frac{1}{1+e^{-(a_i^T x)}}$

$A \quad y$

$$\begin{bmatrix} & & \\ & A & \\ & & \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 1 \\ \vdots \end{bmatrix}$$

← cheap
← expensive
Logistic

Model x

a_i^T

$s(a_i^T x) \geq 0.5$

a_i^T belongs to class 1

≤ 0.5

a_i^T belongs to class 0

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

real number

$a_i^T x$ $\quad a_i^T x$ $\quad s(a_i^T x)$

Linear $\quad a_{n+1}^T x^*$

$x \in \mathbb{R}^d$ $\quad s(a_{n+1}^T x)$ ← Squashing

$= \frac{1}{1+e^{-(a_{n+1}^T x)}}$ $\quad \geq 0.5$, ≤ 0.5

$[0, 1]$ $\quad 0$

$a^T x$

$\sum_{i=1}^{n} (a_i^T b - b_i)^2$ $\quad b_i$'s infi

$-\sum_{i=1}^{n} \left[y_i \ln \pi_i + (1-y_i) \ln(1-\pi_i) \right]$