Logistic Regression

Feb 2023

Sumit Kumar Yadav

Department of Management Studies
Indian Institute of Technology, Roorkee

Recap and Today

- Multicollinearity Issues in Regression
- Logistic Regression
- VIF
- Logistic Regression - More discussion
- Softmax

Example - A customer purchase decision for onion during her visits to a vegetable store are shown below in the table.

Visit Index	1	2	3	4	5	6	7	8	9	10
Price	2	2	2	2	2	3	3	3	3	4
Decision	Y	Y	Y	N	Y	Y	Y	N	N	Y

Visit Index	11	12	13	14	15	16	17	18	19	20
Price	4	4	4	4	4	5	5	5	5	5
Decision	N	N	Y	N	N	N	N	N	N	Y

When the price is 2.5 or when the price is 6, what is the probability that the person will make a purchase?

When the price is 2.5 (or 6), what is the probability that the person will make a purchase?

Price	No. of Visits	No. of Purchases	Probability of Purchase
2	5	4	0.8
3	4	2	0.5
4	6	2	0.33
5	5	1	0.2

Example Contd -

Example - A customer purchase decision for onion during her visits to a vegetable store are shown below in the table.

VI	1	2	3	4	5	6	7	8	9	10
P	2.1	2.05	1.98	2	2.2	3.1	3.06	3.02	3.2	4.1
D	Y	Y	Y	N	Y	Y	Y	N	N	Y

VI	11	12	13	14	15	16	17	18	19	20
P	4.07	4.12	4.14	3.98	4.08	5.1	4.99	5.2	5.15	4.98
D	N	N	Y	N	N	N	N	N	N	Y

When the price is 2.5 or when the price is 6, what is the probability that the person will make a purchase?

Maximum Likelihood Principle - Ideas

Rain Prediction Model				
	Monday	Tuesday	Wednesday	Thursday
Model A	0.3	0.6	0.7	0.2
Model B	0.2	0.4	0.9	0.6
Rained?	N	Y	Y	N

Which Model is better??

Logistic Regression - Main Ideas

Let

$w = [\alpha, \beta_1, \beta_2, ..., \beta_k]$

$x = [1, x_1, x_2, x_3, ..., x_k]$

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

$$z = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

$$P(Yes) = \frac{e^{z}}{1+e^{z}} = \frac{1}{1+e^{-z}}$$

$$P(No) = \frac{1}{1+e^{z}}$$

Logistic Regression - Main Ideas

Let
$w = [\alpha, \beta_1, \beta_2, ..., \beta_k]$
$x = [1, x_1, x_2, x_3, ..., x_k]$

$$\Pr[\textit{Yes} \text{ given } x] = \frac{1}{1 + \exp(-w \cdot x)}$$
$$\Pr[\textit{No} \text{ given } x] = \frac{1}{1 + \exp(+w \cdot x)}$$

Logistic Regression - Main Ideas

Let
$w = [\alpha, \beta_1, \beta_2, ..., \beta_k]$
$x = [1, x_1, x_2, x_3, ..., x_k]$

$$\Pr[\mathit{Yes} \text{ given } x] = \frac{1}{1 + \exp(-w \cdot x)}$$
$$\Pr[\mathit{No} \text{ given } x] = \frac{1}{1 + \exp(+w \cdot x)}$$

Replace Yes with +1 and No with -1

Logistic Regression - Main Ideas

Let
$w = [\alpha, \beta_1, \beta_2, ..., \beta_k]$
$x = [1, x_1, x_2, x_3, ..., x_k]$
$y = 1$, if Yes, $y = -1$ if No

Logistic Regression - Main Ideas

x_0	x_1	x_2	x_3	D
1	–	–	–	+1
1	–	–	–	-1
1	–	–	–	+1
	–	–	–	+1
				-1

Let
$w = [\alpha, \beta_1, \beta_2, ..., \beta_k]$
$x = [1, x_1, x_2, x_3, ..., x_k]$
$y = 1$, if Yes, $y = -1$ if No

$$\Pr[+1 \text{ given } x] = \frac{1}{1 + \exp(-w \cdot x)}$$

$$\Pr[-1 \text{ given } x] = \frac{1}{1 + \exp(+w \cdot x)}$$

Or more compactly

$$\Pr[y \text{ given } x] = \frac{1}{1 + \exp(-y \times w \cdot x)}$$

Probability of observing the dataset as per the model is

$$\prod_{i=1}^{n} \Pr\left[y_i \text{ given } x_i\right] = \prod_{i=1}^{n} \frac{1}{1+\exp\left(-y_i w \cdot x_i\right)}$$

We are looking for w which will maximize this, or alternatively w which will minimize the expression below -

$$\min_w \sum_{i=1}^{n} \log\left(1+\exp\left(-y_i w \cdot x_i\right)\right)$$

or

$$\min \frac{\left(\sum_{i=1}^{n} \log\left(1+e^{-y_i \vec{w} \cdot \vec{x_i}}\right)\right)1}{n}$$

Logistic Regression - Implementation in Python

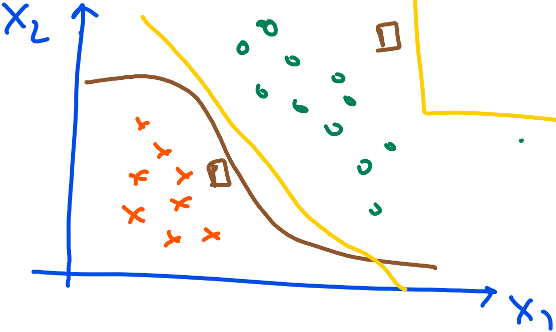
Equivalent Cutoff condition with threshold

α, β_1, β_2

$$P(\text{Green}) = \frac{1}{1+e^{-(\alpha+\beta_1 x_1+\beta_2 x_2)}}$$

$$\frac{1}{1+e^{-(\ )}} > 0.5$$

Let us say we keep the threshold as 0.5. What is the decision surface?



Multi-class classification - Model

$w_1 = [\alpha_1, \beta_{11}, \beta_{12}, \ldots, \beta_{1k}]$

$w_2 = [\alpha_2, \beta_{21}, \beta_{22}, \ldots, \beta_{2k}]$

multiple classes in the data

$\vdots$

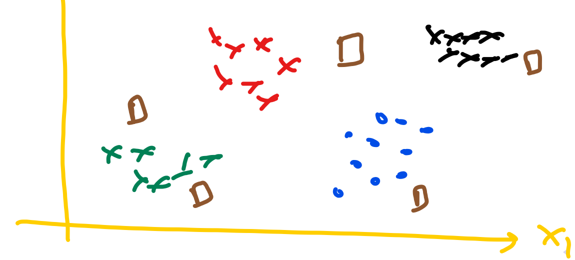
$y \in \{1, 2, 3, \cdots, r\}$

$w_r = [\alpha_r, \beta_{r1}, \beta_{r2}, \ldots, \beta_{rk}]$

Instead of a single weight vector w, we consider r weight vectors

- $w_1, w_2, w_3, \cdots, w_r$.

$$\Pr[\text{ output } i \text{ given } x] = \frac{\exp(w_j \cdot x)}{\sum_{j=1}^{r} \exp(w_j \cdot x)}$$



$[1, 0, -2]$

e^{1}, e^{0}, e^{-2}

$$\frac{e^{1}}{e^{1}+e^{0}+e^{-2}}$$

multiple classes in the data
$y \in \{1, 2, 3, \cdots, r\}$
Instead of a single weight vector w, we consider r weight vectors $w_1, w_2, w_3, \cdots, w_r$.

$$\Pr[\text{ output } i \text{ given } x] = \frac{\exp\left(w_j \cdot x\right)}{\sum_{j=1}^{r} \exp\left(w_j \cdot x\right)}$$

This is called the Soft-Max function, it converts a given set of numbers to probabilities.

X_1	X_2	Y
10	11	A
12	13	B
14	15	C
16	17	B
18	19	A

$\alpha_A, \beta_{A1}, \beta_{A2}$

$\alpha_B, \beta_{B1}, \beta_{B2}$

$\alpha_C, \beta_{C1}, \beta_{C2}$

$$\boxed{O} = \alpha_A + 10 \cdot \beta_{A1} + 11\beta_{A2}$$

$$\boxed{X} = \alpha_B + 10\beta_{B1} + 11\beta_{B2}$$

$$\boxed{+} = \alpha_C + 10\beta_{C1} + 11\beta_{C2}$$

$$\left(\frac{e^{\boxed{O}}}{e^{\boxed{O}} + e^{\boxed{X}} + e^{\boxed{+}}}\right)$$

Softmax for 2 classes

Thank you for your attention