Recap

30 June 2023 19:26

Logistic Reg

linear model

binary classification

Multiclass classification

very technical

Interview

Deep Learning

repeat

What is Multiclass Classification

30 June 2023 14:16

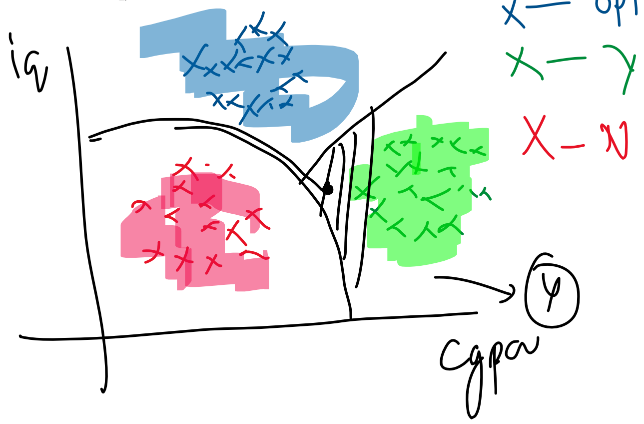
D

cgpa	iq	placement
9	90	Y
6	60	N
7	70	Opt

→ 3 classes

$\boxed{6.5 \quad 65}$

n classes



How to Logistic Regression handles Multiclass Classification Problems

30 June 2023 17:12

binary classification

- → OVR (one vs rest) → OVA (one vs all)
- → Multinomial LR → Softmax Reg

→ K models → K prob → normalize

OVR Approach

30 June 2023 17:13

train → prediction → code

cgpa	iq	placement
–	–	Y
=	–	N
–	–	O

(Y, N, O) → logistic reg model

cgpa	iq	placement = Y	pla = N	pla = O
–	–	1	0	0
–	–	0	1	0
–	–	0	0	1

multiclass (3) → 3 (binary)

binary

cgpa | iq | pla = Y (1, 0) → Logistic Reg → $\beta_0\ \beta_1\ \beta_2$

cgpa | iq | pla = N (1, 0) → Logistic Reg $\beta_0 \beta_1 \beta_2$

cgpa | iq | pla = O → Logi Reg $\beta_0 \beta_1 \beta_2$

model train

Prediction {6.5, 65} → Y, N, O

p(Y) $\boxed{0.6}$ p(N) $\boxed{0.3}$ p(O) $\boxed{0.5}$

$\boxed{1}$ $\boxed{\text{OVR}}$

Normalize → $p(Y) = \frac{0.6}{0.6+0.3+0.5} = \boxed{0.42}$ $p(O) = \frac{0.5}{0.6+0.3+0.5} = \boxed{0.35}$

$p(N) = \frac{0.3}{0.6+0.3+0.5} = \boxed{0.21}$

large → $\boxed{K\uparrow}$ → $\boxed{15 \quad 20}$ → efficient → large dataset

n =

hai high number

$n =$

* efficient → large amount has high number of classes

Code

30 June 2023 17:13

SoftMax Function

30 June 2023 18:00

softmax LR

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

Sigmoid function

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

← binary

prob

$\boxed{0-1}$

z_1 z_2 z_3

$$\sigma(z_1) = \frac{e^{z_1}}{e^{z_1} + e^{z_2} + e^{z_3}}$$

$$\sigma(z_1) + \sigma(z_2) + \sigma(z_3) = 1$$

multiclass

$$\sigma(z_2) = \frac{e^{z_2}}{e^{z_1} + e^{z_2} + e^{z_3}}$$

$$\sigma(z_3) = \frac{e^{z_3}}{e^{z_1} + e^{z_2} + e^{z_3}}$$

SoftMax Logistic Regression / Multinomial LR → multi class

30 June 2023 17:13

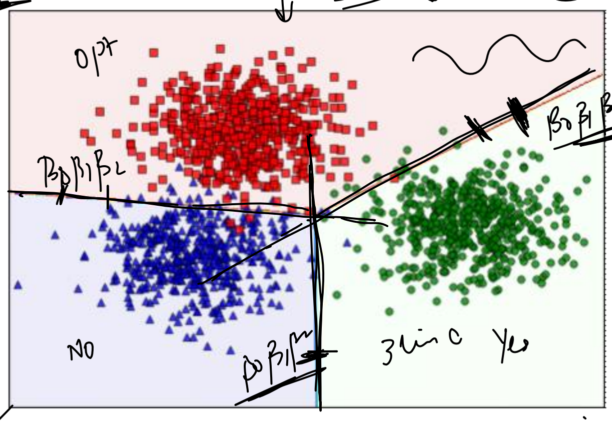
9 parameters

(4) x3 = (12)

|Y, N, O|

4 classes ↓ 4 lines 4x3

cgpa



iq

Logistic → linear model

1 line

$Ax+By+C=0$

$\beta_0+\beta_1x_1+\beta_2x_2=0$

$\boxed{\beta_0 \ \beta_1 \ \beta_2}$

3x3

K classes

↳ softmax LR

↳ K lines to create the decision reg

2d data

features → (2)

training → prediction → code

→ OHE

cgpa	iq	placement = Y	pla = N	pla = O
–	– Y	1	0	0
–	– N	0	1	0
–	– O	0	0	1

↳ loss function →

binary → $\boxed{L = -\frac{1}{n}\sum_{i=1}^{n} y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)}$ $\boxed{k=2}$ → 3 param

minimize

K=3

$\boxed{L = -\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K} y_i^k \log \hat{y}_i^k}$ → Softmax LR

K=2

minimize → 9 paramet

general

categor

special

3 class, 2 input

K = # classes (Y, N, O)

classes → 2 → (K)

→ $\sum_{k=1}^{K} y^k \log \hat{y}^k = \log \hat{y}_1^{yes} +$

$y^{yes}\log \hat{y}^{yes} + \boxed{y^{No}\log \hat{y}^{No}} + \boxed{y^{opt}\log \hat{y}^{opt}}$

cgpa	iq	plac = Y	plac = N	pla = O
–	–	1	0	0
–	–	0	1	0
–	–	0	0	1

n = 1 → i = 1

3 row / 1000

$\log(\hat{y}^{yes}) + 0 + 0$

$\log \hat{y}_2^{No}$

$\log \hat{y}_3^{opt}$

$\log \hat{y}_3^{opt}$

$$-\frac{1}{n}\sum_{i=1}^{n}\sum_{K=1}^{K} y_i^K \log(\hat{y}_i^K)$$

$\boxed{\beta_0^1 \quad \beta_1^1 \quad \beta_2^1}$

output softmax

row number

y_value

$$\boxed{L = \log(\hat{y}_1^{Y}) + \log(\hat{y}_2^{N}) + \log(\hat{y}_3^{opt})}$$

$\beta_0\beta_1\beta_2$ $\quad$ $\beta_0\beta_1\beta_2$ $\quad$ $\beta_0\beta_1\beta_2$

$$\begin{bmatrix} \beta_0^1 & \beta_1^1 & \beta_2^1 \\ \beta_0^2 & \beta_1^2 & \beta_2^2 \\ \beta_0^3 & \beta_1^3 & \beta_2^3 \end{bmatrix}$$

3 random lines

random

min

minimize

$\boxed{\hat{y}_1^{yes}}$, $\hat{y}_2^{no}$, $\hat{y}_3^{opt}$

sigmoid

$\boxed{0, 1}$

$-y_i \log(\hat{y}_i)$

$\boxed{\hat{y}_i}$ → output of logistic

$\boxed{\hat{y}_i} = \sigma(z_i)$

$$\boxed{z_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}}$$

$$\hat{y}_1^{yes} = \sigma(z_i) = \frac{e^{\boxed{z_{yes}}}}{e^{z_{yes}} + e^{z_{no}} + e^{z_{opt}}}$$

softmax

$$\hat{y}_2^{no} = \frac{e^{z_{no}}}{}$$

$$\hat{y}_3^{opt} = \frac{e^{z_{opt}}}{}$$

8 $\quad$ 80

$$z_{yes} = \beta_0 + \beta_1 8 + \beta_2 80$$

8 $\quad$ 80

$$z_{no} = \beta_0^{(2)} + \beta_1^{(2)} 8 + \beta_2^{(2)} 80$$

$$z_{opt} = \beta^{(3)} + \beta^{(3)} 8 + \beta^{(3)} 80$$

9 such

$$\beta_0^1 = \beta_0^1 - \eta \boxed{\frac{\partial L}{\partial \beta_0^1}}$$

9 different

1, 0

multiple

$\boxed{1000}$

1,0

1000

$$\boxed{\frac{\partial L}{\partial \beta_0^1}} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{K} y_i^k \log(\hat{y}_i^k)$$

$\boxed{K=2} \rightarrow$ binary cross entropy

$$= -\frac{1}{n}\sum_{i=1}^{n} y_i^1 \log(\hat{y}_i^1) + y_i^0 \log(\hat{y}_i^0)$$

$(1-y_i^1)$

$\hat{y}_i^0 = (1-\hat{y}_i^1)$

$\boxed{y_i^1 \quad y_i^0}$

$\overset{0}{x}, \overset{1}{y}$

cgpa | iq | placement

$1 \rightarrow y_i^1 = 1 \quad y_i^0 = 0$

$0 \rightarrow y_i^0 = 1 \quad y_i^1 = 0$

$1 - 1 = 0$

$\hat{y}_i = 0.37 \rightarrow p(1) = 0.37$

$p(0) = 0.63$

$$= -\frac{1}{n}\sum_{i=1}^{n} y_i^1 \log(\hat{y}_i^1) + (1-y_i^1)\log(1-\hat{y}_i^1)$$

$$\boxed{= -\frac{1}{n}\sum_{i=1}^{n} y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)}$$

binary cross

grad des

lin1 $\rightarrow$ $\boxed{\beta_0^1 \ \beta_1^1 \ \beta_2^1}$

lin2 $\boxed{\beta_0^2 \ \beta_1^2 \ \beta_2^2}$

lin3 $\boxed{\beta_0^3 \ \beta_1^3 \ \beta_2^3}$

predict

$\{6.5, 65\}$

Prediction

$\rightarrow z_1 = \beta_0^1 + 6.5\beta_1^1 + 65\beta_2^1$

$\rightarrow z_2 = \beta_0^2 + 6.5\beta_1^2 + 65\beta_2^2$

$\rightarrow z_3 = \beta_0^3 + 6.5\beta_1^3 + 65\beta_2^3$

Softmax $\rightarrow$ $\sigma(z_1) = \dfrac{e^{z_1}}{e^{z_1}+e^{z_2}+e^{z_3}} = 0.3$ $\quad \sigma(z_3) = \dfrac{e^{z_3}}{e^{z_1}+e^{z_2}+e^{z_3}}$ 0.5

$\sigma(z_2) = \dfrac{e^{z_2}}{e^{z_1}+e^{z_2}+e^{z_3}}$ 0.2

Code

30 June 2023 17:13

When to use what?

30 June 2023 17:13

Y → 95%

N → 3%

O → 2%

P(Y) = 0.5

P(N) = 0.3

Use One-vs-Rest (OVR) when:

1. Classes are Non-Mutually Exclusive: OVR is appropriate if an instance can belong to more than one class, as each classifier provides an independent probability for each class.
2. Dealing with Imbalanced Data: OVR might perform better when class distribution is highly imbalanced since each class gets a dedicated model.

Use Multinomial Logistic Regression (SoftMax Regression) when:

1. Computational Efficiency is Required: Softmax Regression is generally more efficient for large datasets and a high number of classes.
2. Classes are Mutually Exclusive: SoftMax Regression is a good choice when each instance can only belong to one class. The SoftMax function provides a set of probabilities that sum to 1, fitting well with mutually exclusive classes.
3. Interpretability is Important: The probabilities output by SoftMax Regression are more interpretable than those from OVR, as they always sum to 1. This can make model predictions easier to explain.

Tasks

30 June 2023 18:19

1. Derive sigmoid from softmax
2. Derive binary cross entropy from categorical cross entropy
3. Find the derivative of Softmax Function
4. Find the gradients of cross entropy error

$$\sigma'(z) = \sigma(z)(1-\sigma(z))$$

SVM

DT

RF

Adaboost

GBDT

Xgboost

Kmeans

DBScan

tsne