* Multiclass classification

binary — 1, 0

multiclass classification — 1, 2, 3

iris - 3

Student — Scient, Arts, Commerce

Solution

- → OVR (One Vs rest)
- → multinomial

① One Vs Rest

till now

x_1 - - - x_n → logistic Reg model → Prob value → Threshold

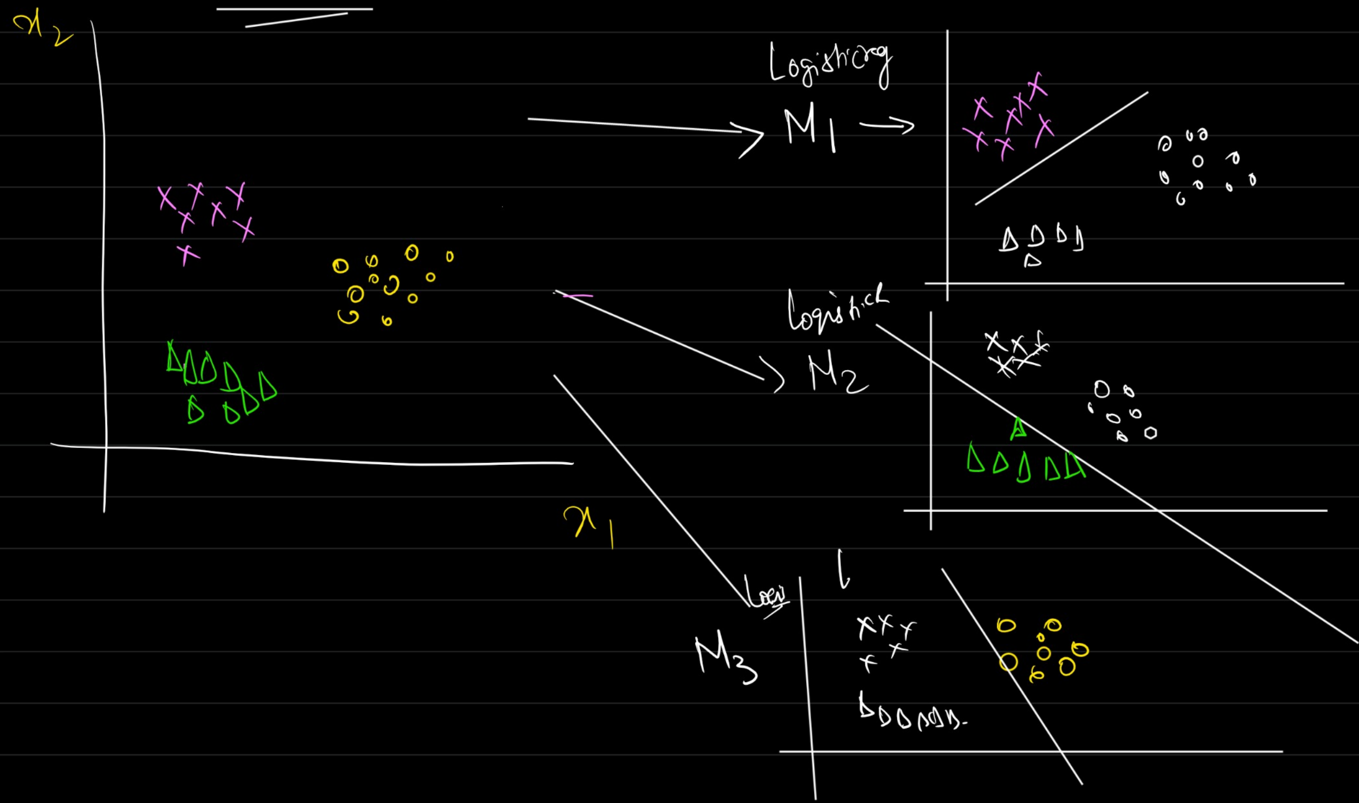
$\geq 0.5 \rightarrow 1$

$< 0.5 \rightarrow 0$

Now

y → 1, 2, 3

One Vs Rest



x_1	x_2	Output	M_1 (1)	M_2 (2)	M_3 (3)
—	—	1	1	0	0
—	—	1	1	0	0
—	—	2	0	1	0
—	—	2	0	1	0
—	—	3	0	0	1
—	—	3	0	0	1
—	—	1	1	0	0
—	—	2	0	1	0
—	—	3	0	0	1

OHE

One Vs Rest

M_1 — Class 1 Vs Rest
(—) (—, —)

M_2 — Class 2 Vs Rest
(—) (—, —)

M_3 Class 3 vs Rest
(—) (—, —)

Expert in class 1 Expert in class 2 Expert class 3

One Vs Rest

x_1	x_2	... x_n	$M_1(1)$	$M_2(2)$	$M_3(3)$	Prediction
—	—	—	0.8	0.6	0.9	Class 3
—	—	—	0.6	0.8	0.5	Class 2

There is 60% chance that this row belongs to class 1

80% chance class 2

50% chance of belong 3

* You will assign the predicted class which gives the highest Probability

Step 1 : # model = no. of Class

Step 2 : Prob of each model

Step 3 : Attach class with highest Prob.

disadvantage → A binary classification model is trained for each class → Computationally expensive.

② Multinomial method / softmax regression

→ We don't decompose the problem into binary classification.

→ Modify the loss / cost fn

→ Single model.

$$\text{Sigmoid} = \frac{1}{1+e^{-z}} = \frac{1}{1+\frac{1}{e^{z}}} = \frac{1}{\frac{e^{z}+1}{e^{z}}}$$

$$= \boxed{\frac{e^{z}}{1+e^{z}}}$$

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

j is no of class

$$\sigma(z)_1 = \frac{e^{z_1}}{e^{z_1}+e^{z_2}+e^{z_3}} \qquad \sigma(z)_2 = \frac{e^{z_2}}{e^{z_1}+e^{z_2}+e^{z_3}} \qquad \sigma(z)_3 = \frac{e^{z_3}}{e^{z_1}+e^{z_2}+e^{z_3}}$$

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

→ multiclass (modification) $\Rightarrow -\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{k} y_k^{(i)} \log \hat{y}_k^{(i)}$

k → no of class

i – no. of dp's

x_1	x_2	y	$y_{k=1}$	$y_{k=2}$	$y_{k=3}$
x_{11}	x_{12}	1	1	0	0
x_{21}	x_{22}	2	0	1	0
x_{31}	x_{32}	3	0	0	1

$y_1^{(1)} \log \hat{y}_1^{(1)} + \underbrace{y_2^{(1)} \log \hat{y}_2^{(1)} + y_3^{(1)} \log \hat{y}_3^{(1)}}_{0}$

$+ \underbrace{y_1^{(2)} \log \hat{y}_1^{(2)}}_{0} + y_2^{(2)} \log \hat{y}_2^{(2)} + \underbrace{y_3^{(2)} \log \hat{y}_3^{(2)}}_{0}$

$+ \underbrace{y_1^{(3)} \log \hat{y}_1^{(3)} + y_2^{(3)} \log \hat{y}_2^{(3)}}_{0} + y_3^{(3)} \log \hat{y}_3^{(3)}$

$$\boxed{CF = y_1^{(1)} \log \hat{y}^{(1)} + y_2^{(2)} \log \hat{y}_2^{(2)} + y_3 \log \hat{y}_3^{(3)}}$$

$\hat{y}_1^{(1)} = \sigma(\theta_0^{(1)} + \theta_1^{(1)} x_{11} + \theta_2^{(1)} x_{12})$ (σ → softmax)

$\hat{y}_2^{(2)} = \sigma(\theta_0^{(2)} + \theta_1^{(2)} x_{21} + \theta_2^{(2)} x_{22})$

$\hat{y}_3^{(3)} = \sigma(\theta_0^{(3)} + \theta_1^{(3)} x_{31} + \theta_3^{(3)} x_{33})$

$\frac{\partial L}{\partial \theta_0}, \frac{\partial L}{\partial \theta_0^{(2)}} \ldots$

→ 9 different

Convergence

$\theta_1^{(1)} = \theta_1^{(1)} - \eta \frac{\partial L}{\partial \theta_1}$

Statmodel implementation Summary explanation

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                 Scores   R-squared:                       0.948
Model:                            OLS   Adj. R-squared:                  0.945
Method:                 Least Squares   F-statistic:                     294.4
Date:                Sun, 19 Mar 2023   Prob (F-statistic):           1.00e-11
Time:                        10:35:25   Log-Likelihood:                -56.886
No. Observations:                  18   AIC:                             117.8
Df Residuals:                      16   BIC:                             119.6
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const          1.9322      3.494      0.553      0.588      -5.476       9.340
Hours          9.9417      0.579     17.158      0.000       8.713      11.170
==============================================================================
Omnibus:                        4.005   Durbin-Watson:                   1.793
Prob(Omnibus):                  0.135   Jarque-Bera (JB):                1.612
Skew:                          -0.338   Prob(JB):                        0.447
Kurtosis:                       1.699   Cond. No.                         15.1
==============================================================================
```

Jarque-Bera (JB): → Normality

Cond. No. → multicon

1. OLS → ordinary least square
2. Dep variable – Score (y)
3. Model → OLS
4. Method – Least Square
5. Date }
6. Time } → when the model was fitted
7. No of obs – the dp used in reg.
8. Df Residuals ⇒ no. of Observation – no of Parameters $= 18 - 2\,(m, c) = 16$
9. df model → parameter − 1 = 2 − 1
10. Rsq – Y explaining
11. Adj Rsquare

12. **F test → Comparing variance**

Hours	Score

$\boxed{y = 0.5x + 2}$ → Hypothesis testing ⇓ my model coeff never become 0.

Comparison → F test

※ fstatistics → measures the overall significance of model. It tests whether atleast one the predictors are non-zero.

$$y_{pr} = \theta_0 + \theta_1 x_1 + \theta_2 x_2 \; ---- \; \theta_n x_n$$

Ftest:

H_0: – All the coeff's are 0, model is insignificant

H_A – Model is significant (atleast one coeff is non zero)

$y = mx + c$

$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 - - \theta_n x_2$

* you have to say if individual features are significant.

$H_0 : \theta_0 = 0$

$H_A : \theta_A \neq 0$