Page
Date / /

2. AUC in classification problems -

Dataset

- Customer ID:
- Gender: ['Male', 'Female']
- Senior Citizen: [1 = Yes, 0 = No]
- Partner: (Yes / No)
- Dependents: (Yes / No)
- Tenure: no. of months customer has stayed with company
- Phone service: (Yes / No)
- Multiple lines: (Yes / No / No phone service)
- Internet service: (DSL, Fiber optic, No)
- Monthly Charges:
- Total charges: incurred by customer.
- Churn: (Yes / No)
- Online security:
- Online backup:
- Device Protection:
- Tech support:
- Streaming TV:
- Streaming movies:
- Contract
- Paperless billing

Page
Date / /

Preprocessing

- Filled missing values of "Total Charges" with median.
- apply one hot encoding to categorical variables

```
"True" : 1
"False" : 0
"Male" : 1
"Female" : 0
"No" : 0
"Yes" : 1
```

Descriptive stats

- Gender: "Male": 3555, "Female": 3488 } almost evenly distributed
- Senior citizen: "No" : 5901, "Yes" : 1142 } most of the dataset doesn't have senior citizen. (potential skewness)
- Partner: "No" : 3641, "Yes" : 3402 } evenly distributed

Page
Date / /

- Phone service:

"No" : 682
"Yes" : 6361

- Multiple lines:

No : 3390
No p.s. : 682
Yes : 2971

- Normalize the numerical features

- tenure
- Monthly charges
- Total charges

using standard normal scaling. (μ(Mean)=0, σ(standard dev.)=1)

Page
Date / /

- Apply binary coding (Yes: 1), (No: 0) into
  - Gender
  - Partner
  - Dependents
  - Phone service
  - Paperless Billing
  - Churn

- One-hot columns (multiple class labels) :
  - Multiple lines
  - Internet Service
  - Online security
  - Online Backup
  - Device Protection
  - Tech Support
  - Streaming Movies
  - Contract
  - Payment Method

Page
Date / /

Correlation analysis

Features with highest correlation are

- Contract_month to month : 0.405
- tenure : 0.352
- Online security No : 0.343
- TechSupport No : 0.337
- Internet Service_Fiberoptic : 0.308
- Contract_Two year : 0.302
- Payment Method : Electronic check : 0.302
- Online Backup_No : 0.268

for simplicity of model we will take 'tenure' and 'total charges'

Now for simplicity we take 5 samples from the dataset.

tenure (x_1)	TotalCharges (x_2)	Churn (y)
1	29.85	0
34	1889.50	0
2	108.15	1
45	1840.75	0
2	151.65	1

x_1 = Tenure
x_2 = Total Charges
y = Churn

Logistic Regression Model

$$\hat{y} = \frac{1}{1+e^{-z}}, \quad z = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Initializing the parameters at zero:

$$\beta_0 = 0, \quad \beta_1 = 0, \quad \beta_2 = 0$$

$$\hat{y}(i) = 0.5$$

Cost Function and Gradients

$$\frac{\partial J}{\partial \beta_j} = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}^{(i)} - y^{(i)}) x_j^{(i)}$$

$$n = 5$$

Page
Date / /

Calculating error term: $e_j = \hat{y} - y$

$e_1 = 0.5 - 0 = 0.5$
$e_2 = 0.5 - 0 = 0.5$
$e_3 = 0.5 - 1 = -0.5$
$e_4 = 0.5 - 0 = 0.5$
$e_5 = 0.5 - 1 = -0.5$

Calculate gradient

- β_0 :

$$\frac{\partial J}{\partial \beta_0} = \frac{1}{5}(0.5 + 0.5 - 0.5 + 0.5 - 0.5) = \frac{0.5}{5} = 0.1$$

- β_1 :

$$\frac{\partial J}{\partial \beta_1} = \frac{38}{5} = 7.6$$

- β_2 :

$$\frac{\partial J}{\partial \beta_2} = \frac{1750.15}{5} = 350.03$$

3. Parameter update

- β_0:

$$\beta_0 \leftarrow \beta_0 - \alpha \frac{\partial J}{\partial \beta_0} = -0.0001$$

- β_1:

$$\beta_1 \leftarrow \beta_1 - \alpha \frac{\partial J}{\partial \beta_1} = -0.0076$$

- β_2:

$$\beta_2 \leftarrow \beta_2 - \alpha \frac{\partial J}{\partial \beta_2} = -0.35003$$

4. Final Model

$$z = -0.0001 - 0.0076x_1 - 0.35003x_2$$

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

Page
Date / /

- sample 0: $x_1 = 1, x_2 = 29.85$

$$z_0 = -0.0001 - 0.0076(1) - 0.33003(29.85)$$

$$\hat{y}_0 = \frac{1}{1 + e^{10.4555}} \approx 2.88 \times 10^{-5}$$

- sample 1: $x_1 = 34, x_2 = 1889.50$

$$\hat{y}_1 = \frac{1}{1 + e^{661.64}} \approx 0$$

- sample 2: $x_1 = 2, x_2 = 108.15$

$$\hat{y}_2 = \frac{1}{1 + e^{37.8710}} \approx 0$$

- sample 3: $x_1 = 45, x_2 = 1840.75$

$$\hat{y}_3 \approx 0$$

- sample 4: $x_2 = 2, x_2 = 151.65$

$$\hat{y}_4 \approx 0$$

Page
Date / /

2: Compute errors $\hat{y}^{(i)} - y^{(i)}$

$e_0 = 2.88 \times 10^{-5}$
$e_1 = 0$
$e_2 = -1$
$e_3 = 0$
$e_4 = -1$

3. Compute Gradients

- $\frac{\partial J}{\partial \beta_0} = -0.399994$

- $\frac{\partial J}{\partial \beta_1} = -0.799994$

- $\frac{\partial J}{\partial \beta_2} = -51.95983$

4. Parameter Update

- $\beta_0 = 0.000300$
- $\beta_1 = -0.006800$
- $\beta_2 = -0.29807$

- model will converge better with more dataset and more iterations.

Final model:

$\beta_0 = 0.0003$
$\beta_1 = -0.0068$
$\beta_2 = -.29807$

$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

$= 0.0003 - 0.0068$ (tenure) $- 0.29807$ (Total charges)

Predicting Churn Probability:

$$P(y=1|x) = \hat{y} = \frac{1}{1+e^{-z}} = \frac{1}{1+e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2)}}$$

- $\hat{y} \geq 0.5$ $(y=1)$ (Churn)
- $\hat{y} < 0.5$ $(y=0)$ (No Churn)

Page
Date / /

- Test 1: (Tenure: 1, Total Charges: 29.85)

$$P(y=1|x) = \frac{1}{1+e^{-(-8.9862)}} = \frac{1}{1+7369.21}$$

$$\approx 0.0001357$$

Predicted Churn: No

- Test 2: (Tenure: 34, Total Charges: 1889.5)

$$P(y=1|x) = \frac{1}{1+e^{563.3975}} \approx 0$$

Predicted Churn: No

(Model is correct for both cases.)

- Test 3: (Tenure: 71, Total Charges: 7382.25)

$$P(y=1|x) = \frac{1}{1+e^{2200.46}} \approx 0$$

Predicted Churn: No

Page
Date / /

• Test 4: (Tenure : 10, Total Charges : 528.35)

$$P(y=1|x) \approx \frac{1}{1+e^{157.78}} \simeq 0$$

Prediction Churn : No (actual prediction : Yes)

• Test 5: (Tenure : 21, Total Charges : 1862.9)

$P(y=1|x) \approx 0$

Prediction Churn : No

• Test 6: (Tenure : 1, Total Charges : 39.65)

$P(y=1|x) \approx 0.0000072$

Prediction Churn : No (actual prediction : Yes)

Page
Date / /

- Test 7: (Tenure: 12, TotalCharges: 202.25)

$P(y=1|x) \approx 0$

Prediction churn: No

- T.N. = 5
- T.P. = 0
- F.P. = 0
- F.N. = 2

$$-TPR = \frac{0}{5+0} = 0$$

$$-FPR = \frac{0}{0+2} = 0$$

∵ both are 0

$\Rightarrow AUC = 0$

- model is not good it is not applicable but with more training data and evaluation (testing) data it would be better.