



Osztályozás Metrikák

Data science képzés

Log-loss

2024.10.07.

Jónás Dániel,
data scientist

A film vígjáték-e?

y	y_pred_proba	y_pred
0	0.5	1
1	0.9	1
0	0.7	1
1	0.7	1
1	0.3	0
0	0.4	0
1	0.5	1

		Aktuális (y)	
		Pozitív (1)	Negatív (0)
Prediktált (y_pred)	Pozitív (1)	3	2
	Negatív (0)	1	1

A film vígjáték-e?

$$\text{Precision} = 3 / (3 + 2) = 3/5$$

y	y_pred_proba	y_pred
0	0.5	1
1	0.9	1
0	0.7	1
1	0.7	1
1	0.3	0
0	0.4	0
1	0.5	1

		Aktuális (y)	
		Pozitív (1)	Negatív (0)
Prediktált (y_pred)	Pozitív (1)	3	2
	Negatív (0)	1	1

A film vígjáték-e?

$$\text{Precision} = 3 / (3 + 2) = 3/5$$

$$\text{Recall} = 3 / (3 + 1) = 3/4$$

y	y_pred_proba	y_pred
0	0.5	1
1	0.9	1
0	0.7	1
1	0.7	1
1	0.3	0
0	0.4	0
1	0.5	1

		Aktuális (y)	
		Pozitív (1)	Negatív (0)
Prediktált (y_pred)	Pozitív (1)	3	2
	Negatív (0)	1	1

A film vígjáték-e?

$$\text{Precision} = 3 / (3 + 2) = 3/5$$

$$\text{Recall} = 3 / (3 + 1) = 3/4$$

$$\text{Accuracy} = 4 / (3 + 2 + 1 + 1) = 4 / 7$$

y	y_pred_proba	y_pred
0	0.5	1
1	0.9	1
0	0.7	1
1	0.7	1
1	0.3	0
0	0.4	0
1	0.5	1

		Aktuális (y)	
		Pozitív (1)	Negatív (0)
Prediktált (y_pred)	Pozitív (1)	3	2
	Negatív (0)	1	1

A film vígjáték/akció/horror?

Aktuális

	Vígjáték	Akció	Horror
Vígjáték	7	8	9
Akció	1	2	3
Horror	3	2	1

	Vígjáték	Akció	Horror
TP			
TN			
FP			
FN			

Prediktált

A film vígjáték/akció/horror?

Aktuális

	Vígjáték	Akció	Horror
Vígjáték	7	8	9
Akció	1	2	3
Horror	3	2	1

	Vígjáték	Akció	Horror
TP	7		
TN			
FP			
FN			

Prediktált

A film vígjáték/akció/horror?

Aktuális

	Vígjáték	Akció	Horror
Vígjáték	7	8	9
Akció	1	2	3
Horror	3	2	1

	Vígjáték	Akció	Horror
TP	7		
TN	8		
FP			
FN			

Prediktált

A film vígjáték/akció/horror?

Aktuális

	Vígjáték	Akció	Horror
Vígjáték	7	8	9
Akció	1	2	3
Horror	3	2	1

	Vígjáték	Akció	Horror
TP	7		
TN	8		
FP	17		
FN			

Prediktált

A film vígjáték/akció/horror?

Prediktált

Aktuális

	Vígjáték	Akció	Horror
Vígjáték	7	8	9
Akció	1	2	3
Horror	3	2	1

	Vígjáték	Akció	Horror
TP	7		
TN	8		
FP	17		
FN	4		

A film vígjáték/akció/horror?

Aktuális

	Vígjáték	Akció	Horror
Vígjáték	7	8	9
Akció	1	2	3
Horror	3	2	1

	Vígjáték	Akció	Horror
TP	7	2	
TN	8		
FP	17		
FN	4		

Prediktált

A film vígjáték/akció/horror?

Prediktált

Aktuális

	Vígjáték	Akció	Horror
Vígjáték	7	8	9
Akció	1	2	3
Horror	3	2	1

	Vígjáték	Akció	Horror
TP	7	2	
TN	8	20	
FP	17		
FN	4		

A film vígjáték/akció/horror?

Prediktált

Aktuális

	Vígjáték	Akció	Horror
Vígjáték	7	8	9
Akció	1	2	3
Horror	3	2	1

	Vígjáték	Akció	Horror
TP	7	2	
TN	8	20	
FP	17	4	
FN	4		

A film vígjáték/akció/horror?

Aktuális

	Vígjáték	Akció	Horror
Vígjáték	7	8	9
Akció	1	2	3
Horror	3	2	1

	Vígjáték	Akció	Horror
TP	7	2	
TN	8	20	
FP	17	4	
FN	4	10	

Prediktált

A film vígjáték/akció/horror?

Aktuális

	Vígjáték	Akció	Horror
Vígjáték	7	8	9
Akció	1	2	3
Horror	3	2	1

	Vígjáték	Akció	Horror
TP	7	2	1
TN	8	20	18
FP	17	4	5
FN	4	10	12

Prediktált

A film vígjáték/akció/horror?

	Vígjáték	Akció	Horror
TP	7	2	1
TN	8	20	18
FP	17	4	5
FN	4	10	12

A film vígjáték/akció/horror?

	Vígjáték	Akció	Horror
TP	7	2	1
TN	8	20	18
FP	17	4	5
FN	4	10	12

	Precision TP / (TP + FP)	Recall TP / (TP + FN)
Vígjáték	7 / 24	7 / 11
Akció	2 / 6	2 / 12
Horror	1 / 6	1 / 13

A film vígjáték/akció/horror?

	Vígjáték	Akció	Horror	Egész
TP	7	2	1	10
TN	8	20	18	46
FP	17	4	5	26
FN	4	10	12	26

	Precision TP / (TP + FP)	Recall TP / (TP + FN)
Vígjáték	7 / 24	7 / 11
Akció	2 / 6	2 / 12
Horror	1 / 6	1 / 13
TOTAL	10/36	10/36

A film vígjáték/akció/horror?

	Vígjáték	Akció	Horror	Egész
TP	7	2	1	10
TN	8	20	18	46
FP	17	4	5	26
FN	4	10	12	26

	Precision TP / (TP + FP)	Recall TP / (TP + FN)
Vígjáték	0.29	0.63
Akció	0.33	0.16
Horror	0.16	0.08
TOTAL	0.28	0.28

Egy nagy metrika

Log loss

- Korábbi metrikákkal ellentétben, ez az osztálytól független érték
- $\text{log loss} = -1 * (\log(\text{likelihood function}) / n)$
- Minél magabiztosabban találjuk el az eredményt, annál jobb
- Minél magabiztosabban lövünk mellé, annál jobban büntet
- Osztályok számától és bennük előforduló értékek számától függően értelmezendő

Likelihood function

- aktuális értékek = $[1, 0, 0]$
- prediktált valószínűségek = $[0.8, 0.3, 0.1]$

Likelihood function

- aktuális értékek = $[1, 0, 0]$
- prediktált valószínűségek = $[0.8, 0.3, 0.1]$
- $1: \text{akt} == 1, P(1) \rightarrow 0.8$

Likelihood function

- aktuális értékek = $[1, 0, 0]$
- prediktált valószínűségek = $[0.8, 0.3, 0.1]$
- 1: $\text{akt} == 1, P(1) \rightarrow 0.8$
- 2. $\text{akt} == 0, P(0) \rightarrow 0.7 \text{ (} 1 - 0.3 \text{)}$

Likelihood function

- aktuális értékek = $[1, 0, 0]$
- prediktált valószínűségek = $[0.8, 0.3, 0.1]$
- 1: $\text{akt} == 1, P(1) \rightarrow 0.8$
- 2. $\text{akt} == 0, P(0) \rightarrow 0.7 \text{ (} 1 - 0.3 \text{)}$
- 3. $\text{akt} == 0, P(0) \rightarrow 0.9 \text{ (} 1 - 0.1 \text{)}$

Likelihood function

- aktuális értékek = [1, 0, 0]
- prediktált valószínűségek = [0.8, 0.3, 0.1]
- 1: akt == 1, $P(1) \rightarrow 0.8$
- 2. akt == 0, $P(0) \rightarrow 0.7$ ($1 - 0.3$)
- 3. akt == 0, $P(0) \rightarrow 0.9$ ($1 - 0.1$)
- Log Likelihood function / n $\rightarrow \log(0.8 * 0.7 * 0.9) / 3 = -0.228$

Likelihood function

- aktuális értékek = [1, 0, 0]
- prediktált valószínűségek = [0.8, 0.3, 0.1]
- 1: akt == 1, $P(1) \rightarrow 0.8$
- 2. akt == 0, $P(0) \rightarrow 0.7$ ($1 - 0.3$)
- 3. akt == 0, $P(0) \rightarrow 0.9$ ($1 - 0.1$)
- Log Likelihood function / n $\rightarrow \log(0.8 * 0.7 * 0.9) / 3 = -0.228$
- log loss = $-1 * \log(0.504) / 3 = -1 * -0.228 = 0.228$

Likelihood function

- aktuális értékek = $[0, 0, 2]$ (lehetséges indexek: 0, 1, 2)
- prediktált valószínűségek ==
- $[[0.6, 0.3, 0.1],$
- $[0.3, 0.5, 0.2],$
- $[0.1, 0.4, 0.5]]$

Likelihood function

- aktuális értékek = $[0, 0, 2]$ (lehetséges indexek: 0, 1, 2)
- prediktált valószínűségek ==
- $[[0.6, 0.3, 0.1],$
- $[0.3, 0.5, 0.2],$
- $[0.1, 0.4, 0.5]]$
- 1: akt == 0, $P(0) \rightarrow 0.6$

Likelihood function

- aktuális értékek = $[0, 0, 2]$ (lehetséges indexek: 0, 1, 2)
- prediktált valószínűségek ==
- $[[0.6, 0.3, 0.1],$
- $[0.3, 0.5, 0.2],$
- $[0.1, 0.4, 0.5]]$
- 1: akt == 0, $P(0) \rightarrow 0.6$
- 2. akt == 0, $P(0) \rightarrow 0.3$

Likelihood function

- aktuális értékek = $[0, 0, 2]$ (lehetséges indexek: 0, 1, 2)
- prediktált valószínűségek ==
- $[0.6, 0.3, 0.1],$
- $[0.3, 0.5, 0.2],$
- $[0.1, 0.4, 0.5]$
- 1: $\text{akt} == 0, P(0) \rightarrow 0.6$
- 2. $\text{akt} == 0, P(0) \rightarrow 0.3$
- 3. $\text{akt} == 2, P(2) \rightarrow 0.5$

Likelihood function

- aktuális értékek = [0, 0, 2] (lehetséges indexek: 0, 1, 2)
- prediktált valószínűségek ==
- [[0.6, 0.3, 0.1],
- [0.3, 0.5, 0.2],
- [0.1, 0.4, 0.5]]
- 1: akt == 0, $P(0) \rightarrow 0.6$
- 2. akt == 0, $P(0) \rightarrow 0.3$
- 3. akt == 2, $P(2) \rightarrow 0.5$
- Log probability function / $n \rightarrow \log(0.6 * 0.3 * 0.5) / 3 = -0.803$

Likelihood function

- aktuális értékek = [0, 0, 2] (lehetséges indexek: 0, 1, 2)
- prediktált valószínűségek ==
- [[0.6, 0.3, 0.1],
- [0.3, 0.5, 0.2],
- [0.1, 0.4, 0.5]]
- 1: akt == 0, $P(0) \rightarrow 0.6$
- 2. akt == 0, $P(0) \rightarrow 0.3$
- 3. akt == 2, $P(2) \rightarrow 0.5$
- Log probability function / $n \rightarrow \log(0.6 * 0.3 * 0.5) / 3 = -0.803$
- log loss = $-1 * \log(0.09) / 3 = -1 * -0.803 = 0.803$

Ideális log loss

- aktuális értékek = $[0, 0, 2]$ (lehetséges indexek: 0, 1, 2)
- prediktált valószínűségek ==
- $[[1, 0, 0],$
- $[1, 0, 0],$
- $[0, 0, 1]]$
- 1: akt == 0, $P(0) \rightarrow 1$
- 2. akt == 0, $P(0) \rightarrow 1$
- 3. akt == 2, $P(2) = 1$
- Log probability function / $n \rightarrow \log(1 * 1 * 1) / 3 = 0$
- log loss = $-1 * \log(1) / 3 = -1 * 0 = 0$

Egy könnyen értelmezhető metrika

Üzleti metrikák

- Az üzleti hatás nehezen értelmezhető a korábbi metrikákkal
- Kreatív módon kézzel készíthető metrikák
- Ezekre a modell nem tanítható (vagy nem ideálisak rá), de emberi fogyasztásra alkalmasak

Egy könnyen értelmezhető metrika

Üzleti metrikák - példa

	Vígjáték	Akción	Horror	Akt
1	0.6	0.3	0.1	Vígjáték
2	0.1	0.85	0.05	Akción
3	0.5	0.2	0.3	Akción
4	0.7	0.13	0.17	Horror

Egy könnyen értelmezhető metrika

Üzleti metrikák - példa

	Vígjáték	Akción	Horror	Akt	Pred Sorrend
1	0.6	0.3	0.1	Vígjáték	V, A, H
2	0.1	0.85	0.05	Akción	A, V, H
3	0.5	0.2	0.3	Akción	V, H, A
4	0.7	0.13	0.17	Horror	V, H, A

Egy könnyen értelmezhető metrika

Üzleti metrikák - példa

	Vígjáték	Akción	Horror	Akt	Pred Sorrend	Rank
1	0.6	0.3	0.1	Vígjáték	V, A, H	1
2	0.1	0.85	0.05	Akción	A, V, H	1
3	0.5	0.2	0.3	Akción	V, H, A	3
4	0.7	0.13	0.17	Horror	V, H, A	2