

Session 11

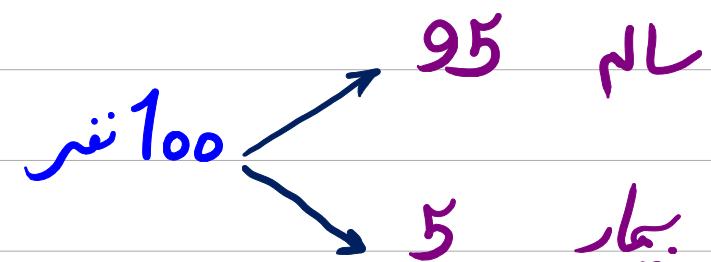
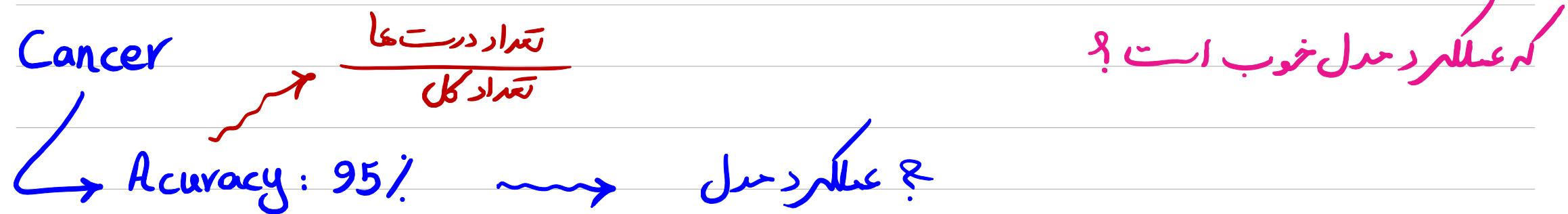
F1-Score & One-Hot-Encoding

Machine Learning | Hobot | Zahra Amini

Telegram: @zahraamini_ai & Instagram:@zahraamini_ai & LinkedIn: @zahraamini-ai

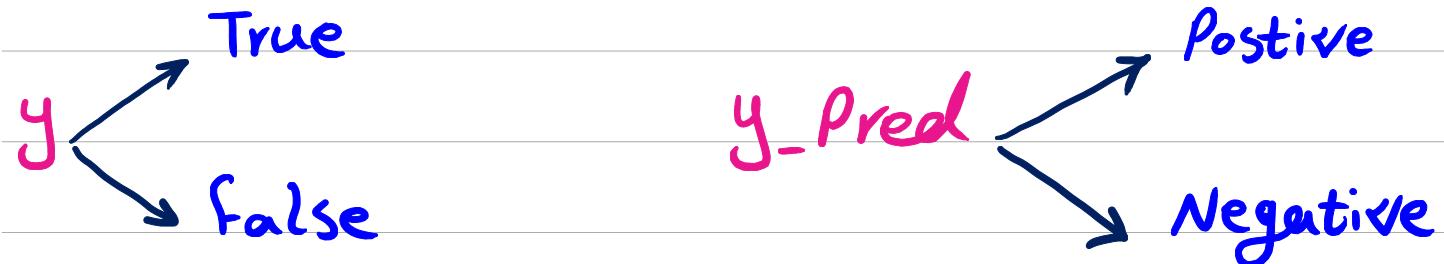
<https://zil.ink/zahraamini>

؟ آندر مسائل classification فقط از accuracy برای ارزیابی مدل استفاده کنیم که یا حتی توافق نهیم. لیکن



مدل ما آن رهیشه "سالم" را پیش‌بینی کند با
دقت 95٪ دارد درست عمل کند. :-

برای رفع این خشکی باید از Confusion Matrix استفاده کنیم.



True Positive \rightarrow Predict Positive , True

True Negative \rightarrow Predict Negative , True

False Positive \rightarrow Predict Positive, False

False Negative \rightarrow Predict Negative, False

TP TN

FP FN

y		$\text{thr} = 0.6$	Recall	Precision	Accuracy
$TN \leftarrow$	0	0.5	0		
$TP \leftarrow$	1	0.9	1		
$FP \leftarrow$	0	0.7	1		
$TP \leftarrow$	1	0.7	1	$\frac{1}{2}$	$\frac{2}{3}$
$FN \leftarrow$	1	0.3	0		
$TN \leftarrow$	0	0.4	0		
$FN \leftarrow$	1	0.5	0		

		Actual	
		1	0
Predict	1	TP	FP
	0	FN	TN

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{Total}}$$

$$\text{F1-Score} = \frac{2}{\frac{1}{\text{Recall}} + \frac{1}{\text{Precision}}}$$

بیشتر مدل های یادگیری ماشین با داده ها عددی کار می کنند.

ج. الگوریتم بر روی داده های Categorical کار نمی کند، چه طور باید این مشکل را حل کنیم؟

خوب ساده است باید داده های این را به عدد تبدیل کنیم، اما چه طور؟

1. Integer Encoding

2. One-Hot Encoding

"Red": 1

"Yellow": 2

"Green": 3

Color
Red
Red
Yellow
Green
Yellow

Integer Encoding →

Color
1
1
2
3
2

ظاهر آنکه همه چیز به نظر خوب می‌رسد، اما واقعاً اینجوریه؟ ordered relationship

الگریابی هر Category یک عدد بدهیم و مدل ما ارزش آن عدد را در روابط دخیل آن در حالی که

عدد بزرگتر شان دهنده‌ی برتری یک category نسبت به دیگری نیست.

ج: برای رفع این مشکل چه کنم؟ می‌توانیم از One-Hot Encoding استفاده کنیم.

Color	Red	Yellow	Green
Red	1	0	0
Red	1	0	0
Yellow	0	1	0
Green	0	0	1
Yellow	0	1	0

جس سکلت one-Hot Encoding چیت؟

1. افزایش تعداد دیشکیا

Dummy Variable Trap .2

Red	Yellow	Green
1	0	0
1	0	0
0	1	0
0	0	1
0	1	0



Red	Yellow
1	0
1	0
0	1
0	0
0	1