

Session 7&8

F1-Score & One-Hot Encoding

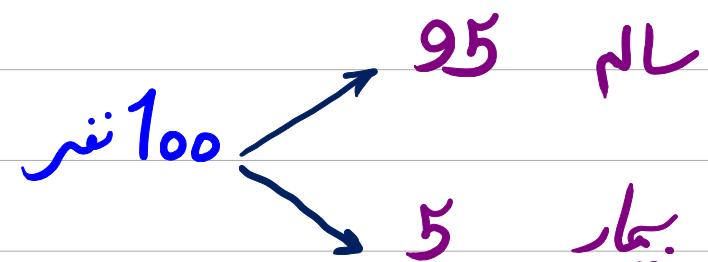
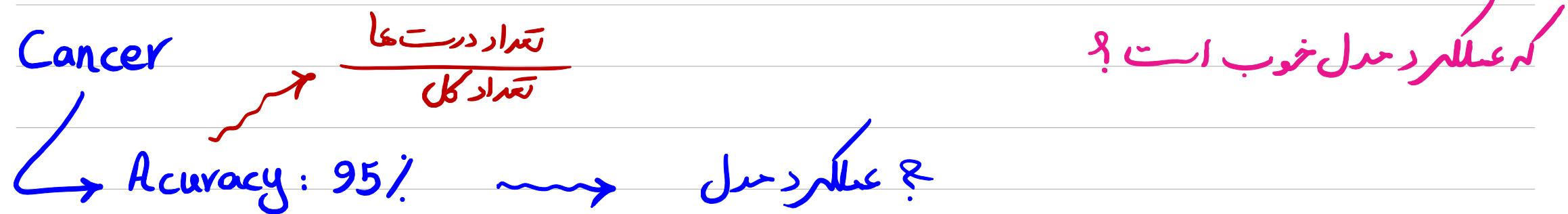
Cost vs. Loss Function

Machine Learning | Zahra Amini

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

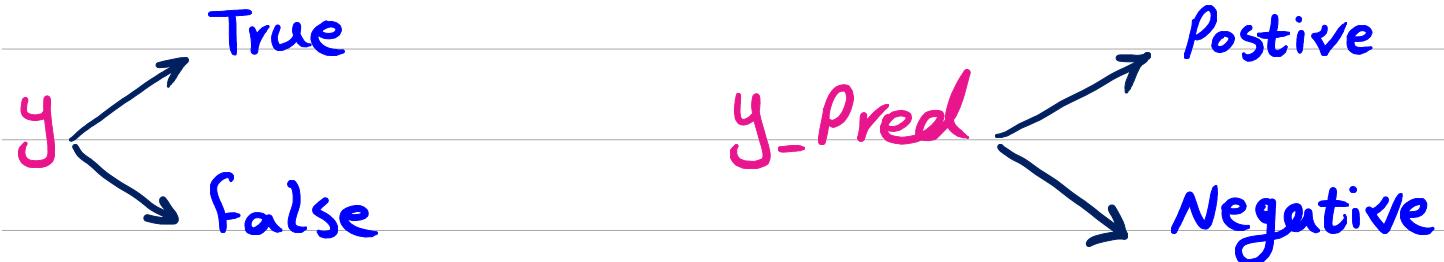
<https://zil.ink/zahraamini>

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



مدل ما آنر هیشت "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 شکل 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 سُلکت ہے؟ one-Hot Encoding

1. افزایش تعداد دیش کی حا

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

Loss vs. Cost Function

تابع خطا Loss Function

میزان خطای یک پیش‌بینی خاص مدل را نسبت به مقدار واقعی محاسبه می‌کند. به عبارت دیگر، این تابع برای یک نمونه منفرد یا داده خاص استفاده می‌شود

$$\text{for MSE} \rightarrow \text{Loss}(y, \hat{y}) = (y - \hat{y})^2$$

تابع هزینه Cost Function

میانگین یا مجموع Loss Function‌ها روی کل نمونه‌های موجود در مجموعه داده است. این تابع عملکرد کلی مدل را ارزیابی می‌کند و مبنای بهینه‌سازی مدل قرار می‌گیرد

$$\text{for MSE} \rightarrow \text{Cost} = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2$$

Cost Function	Loss Function	ویژگی
برای کل مجموعه داده	برای یک نمونه خاص	سطح محاسبه
ارزیابی عملکرد کلی مدل	اندازه‌گیری خطای نمونه منفرد	هدف استفاده
برای آموزش مدل و تنظیم پارامترها	برای محاسبه خطای لحظه‌ای	کاربرد