

INFO370 Lab: confusion matrix

February 26, 2022

Instructions

The goal of this lab is to understand confusion matrix, and the related concepts accuracy, precision and recall.

The lab expects you to compute the confusion matrix *manually*, you may use a calculator for A, P and R. This lab is about thinking and understanding, not coding! Write your answers as valid markdown text, including confusion matrices as markdown tables!

You can create a confusion matrix as a md table like this:

			Predicted			
	-----		-----:		-----:	
			Poisonous		Edible	
	Actual		1		3	
			2		4	

It does not really matter below how do you select positives or negatives. Both approaches are fine, e.g. positive may be edible or poisonous. Just you have to look either at recall or precision, depending on your choice here. Alternatively, you can also talk about e.g. “true edibles” and “false poisonous” instead of positives and negatives.

1 Confusion Matrix

Poisonous mushrooms You get an internship at the mycology department. You are designing an image recognition algorithm for prof. Joffe to distinguish between poisonous (P) and edible (E) mushrooms. You design two models: M_1 and M_2 . The performance on test data is in Table 1, left panel.

1. Show the confusion matrices for models M_1 and M_2 .

NB! *construct these manually and show as markdown table!* Do not create data frames and code! It is easy to find the relevant library routines, but we want you to *understand* what the concepts mean.

2. Compute accuracy, precision, recall for both models.
3. Prof. Joffe wants to get the app out tomorrow. Which model, M_1 or M_2 will you recommend him to use? Explain your reasoning!

Remember: prof Joffe is a mycologist who knows nothing about coding and data science. Explain it in a way he can understand!

Table 1: True value, and model predictions on the test data. M_1 , M_2 and so on denote predictions by the corresponding models.

Mushrooms			Defendants			S. vorax		
Actual	M_1	M_2	Actual	M_3	M_4	Actual	M_5	M_6
P	P	P	G	G	I	S	S	S
E	P	E	I	G	I	O	O	S
E	E	E	I	I	I	S	O	S
P	P	E	G	G	G	S	S	S
P	P	P	I	I	I	O	S	S
E	E	E	G	G	I	S	S	S
E	P	E	G	I	G	O	O	O
E	E	E	I	G	I	O	O	O
E	P	E	I	I	I	O	S	S
E	E	E	G	G	I	S	S	S

Defendants Afterwards you get a summer job in King County Superior Court. [Sheriff Johanknecht](#) wants you to develop an ML algorithm that classifies the defendants as guilty (G) or innocent (I) based on accessible evidence. Although the final decision is done by the judge, your results will weight heavily in that decision. You devise two models, M_3 and M_4 . The test run results are in Table 1, middle panel.

4. Show the confusion matrices for M_3 and M_4 .
5. Compute the accuracy, precision, recall for both models.
6. Johanknecht wants to commission the AI system tomorrow. Which model would you recommend her to use? Explain your reasoning!

Be aware that Johanknecht is a policewoman who knows nothing about data science and coding. Explain it in a way she can understand!

S. vorax Next fall you become an ML expert at the paleontology department for professor Dong. Your task is to devise a model that can distinguish between bones of *simolestes vorax* (S) and other (O) marine dinosaurs (see Figure 1). You have some test data and you develop two models, M_5 and M_6 . The model performance is in Table 1, right panel.



Figure 1: Reconstruction of *Simolestes vorax*. Dmitry Bogdanov, [CC BY-SA 3.0](#), via Wikimedia Commons

7. Show the confusion matrix for M_5 and M_6 .
8. Compute accuracy, precision, recall, and F-score for both models.
9. Prof Dong wants to submit his paper tomorrow. Which of these two models, M_5 or M_6 will you recommend him? Explain to him.

Note: prof Dong is a paleontologist who knows nothing about data science and such. But he wants to hear your suggestions and reasoning!
