







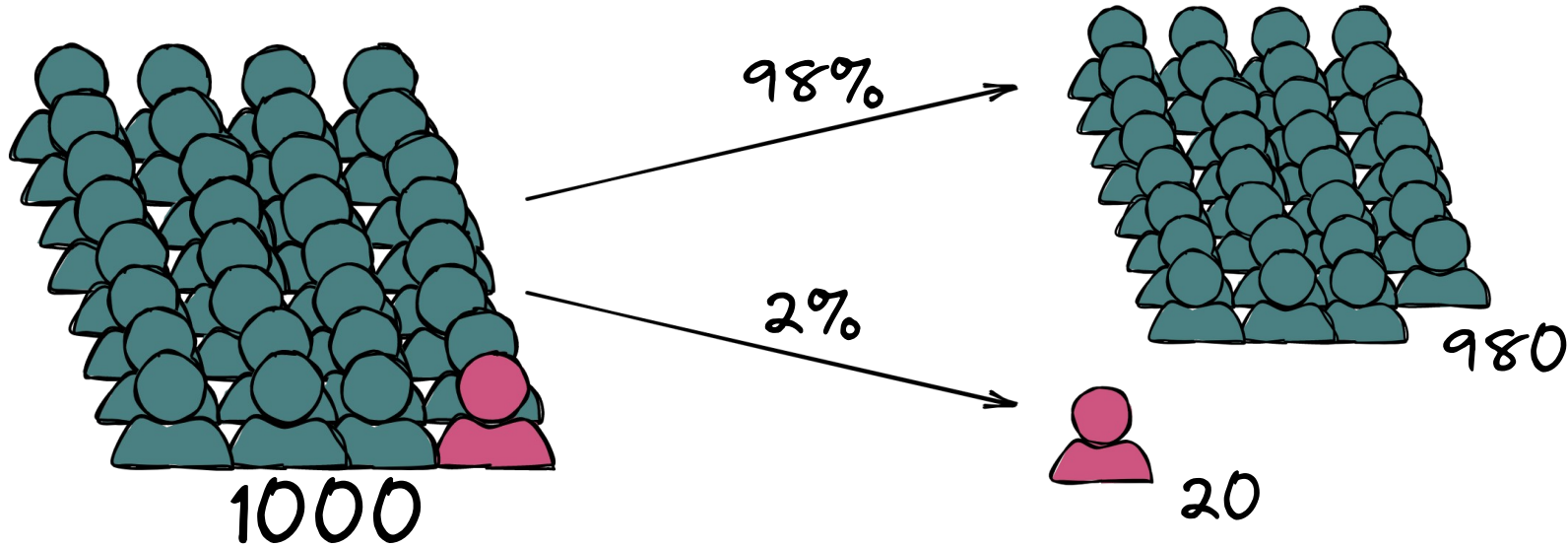


Classification metrics

True Positive act:  pred: 	False Positive act:  pred: 
False Negative act:  pred: 	True Negative act:  pred: 

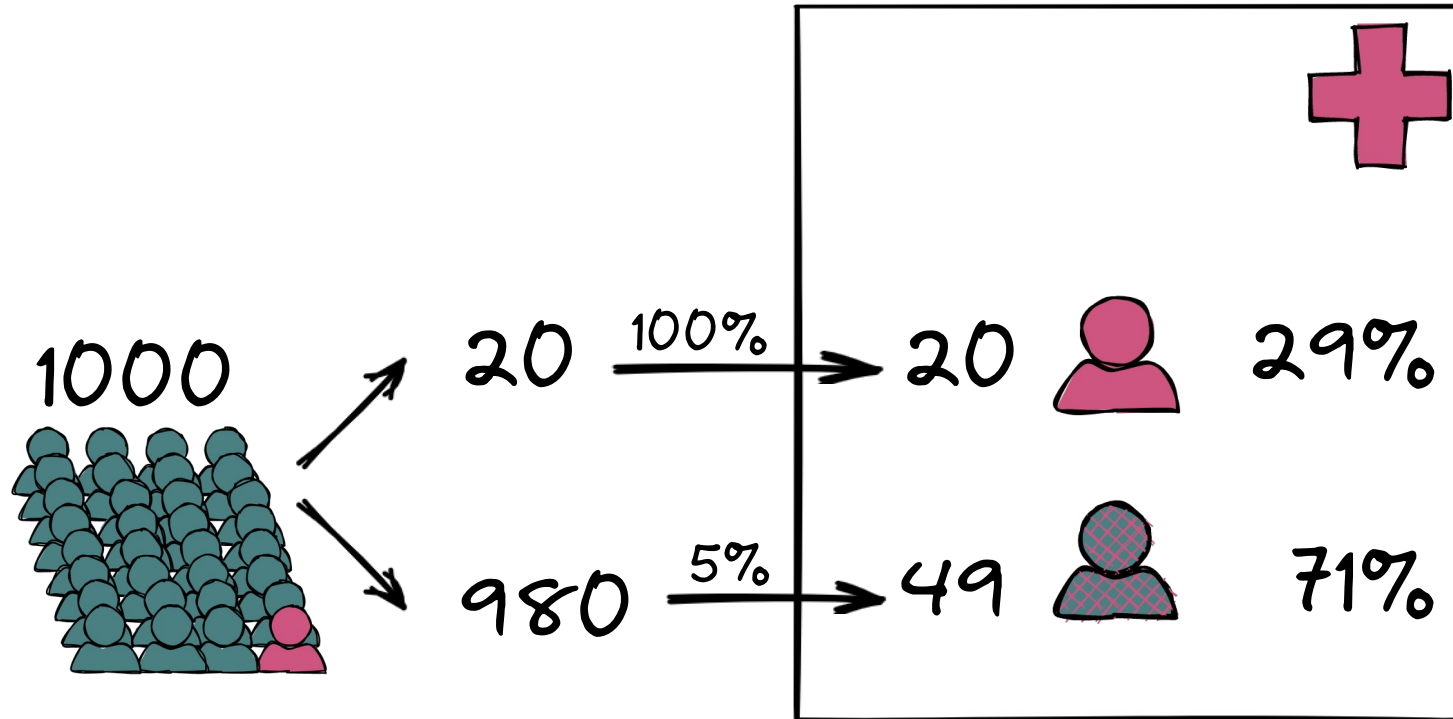
Week 14

Classification metrics: Case1



Disease Test: 100% correct if ill
95% correct if not ill

Classification metrics: Case1











71% of people in the hospital aren't ill.
The error is much higher than 5%

Confusion matrix

actual class: True

actual class: False

predicted: True	<p>True Positive (TP)</p> <p>actual: </p> <p>predicted: </p>	<p>Type I error False Positive (FP)</p> <p>actual: </p> <p>predicted: </p>
predicted: False	<p>Type II error False Negative (FN)</p> <p>actual: </p> <p>predicted: </p>	<p>True Negative (TN)</p> <p>actual: </p> <p>predicted: </p>

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

$$\text{Sensitivity (Recall)} = \frac{TP}{TP+FN}$$

$$\text{Specificity} = \frac{TN}{TN+FP}$$

$$NPV^* = \frac{TN}{TN+FN}$$

$$F1 = 2 \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

*NPV - negative predicted value

Confusion matrix: metrics









Precision relates to the amount of false positive results (type I errors):
high precision means small amount of type I errors.

Recall relates to the amount of false negative results (type II errors):
high recall means small amount of type II errors

One may use the following observation to remember:
precision – positive, recall – negative

F1-score is a metric, taking both precision and recall into account. One of the main reasons, why we multiple precision and recall instead of taking, say, their average, is to deal with extremely low values.

Confusion matrix: Case1

		actual class	
predicted class	True Positive (TP)	False Positive (FP)	
	act:  pred:  931	act:  pred:  0	
False Negative (FN)	True Negative (TN)		
act:  pred:  49	act:  pred:  20		

$$\text{Precision} = \frac{931}{931+0} = 1$$

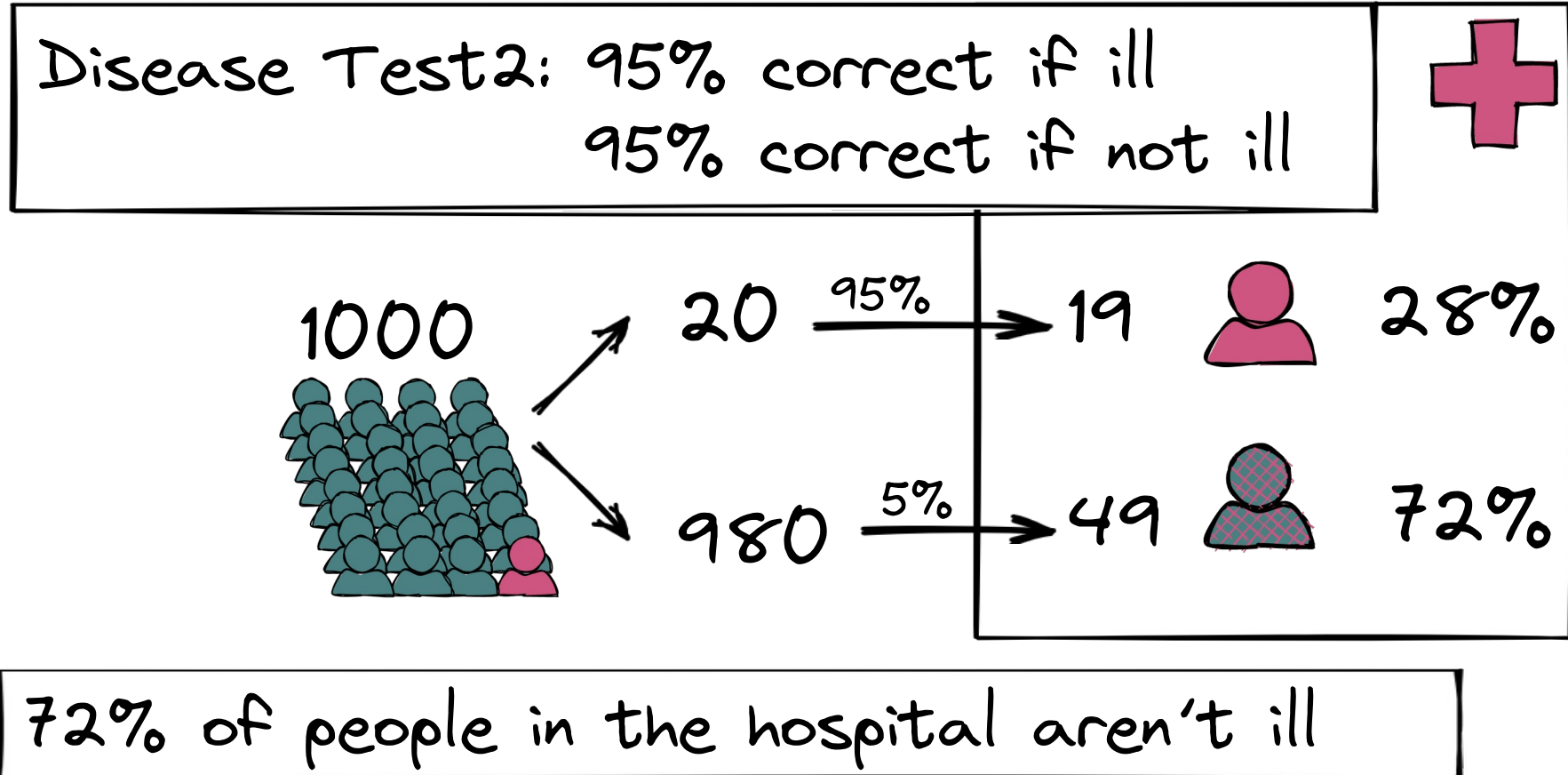
$$\text{Sensitivity (Recall)} = \frac{931}{931+49} = 0.95$$

$$\text{Specificity} = \frac{20}{20+0} = 1$$

$$\text{NPV} = \frac{20}{20+49} = 0.29$$

$$\text{F1} = 2 \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} = 0.97$$

Classification metrics: Case2



Confusion matrix: Case2

Disease Test2: 95% correct if ill
95% correct if not ill

The mentioned TN, TP, FN, FP are easily calculated using Bayes' Theorem:

TP:

$$P(\text{ill if test says ill}) = P(B) * P(A | B) = 0.98 * 0.95 = 0.931 = 93.1\%$$

FN:

$$P(\text{ill if test says not ill}) = 0.98 * 0.05 = 0.049 = 4.9\%$$









TN:

$$P(\text{not ill if test says not ill}) = 0.02 * 0.95 = 0.019 = 1.9\%$$

FP:

$$P(\text{not ill if test says ill}) = 0.02 * 0.05 = 0.001 = 0.1\%$$

Confusion matrix: Case2

		actual class	
predicted class	True Positive (TP)	False Positive (FP)	
	act:  pred:  931	act:  pred:  1	
False Negative (FN)	act:  pred:  49	True Negative (TN)	
		act:  pred:  19	

$$\text{Precision} = \frac{931}{931+1} = 0.99$$

$$\text{Sensitivity (Recall)} = \frac{931}{931+49} = 0.95$$

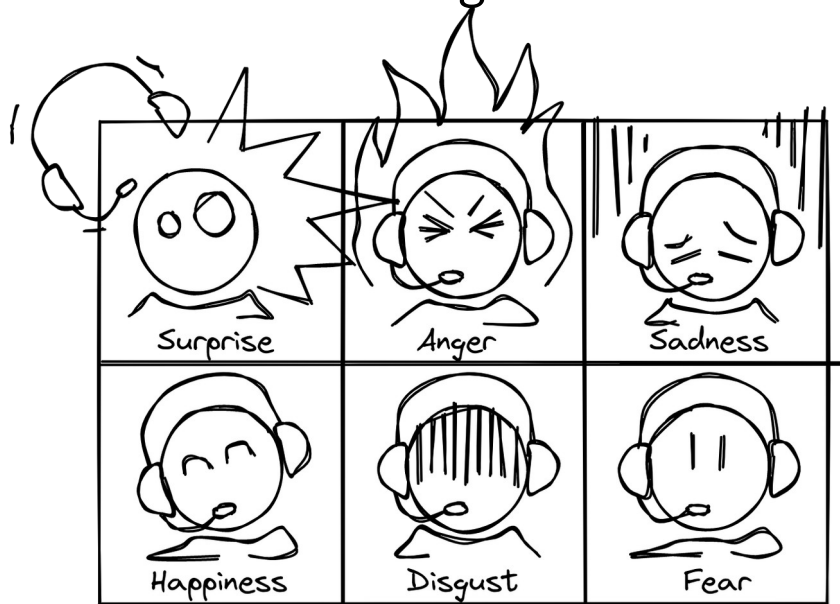
$$\text{Specificity} = \frac{19}{19+1} = 0.95$$

$$\text{NPV} = \frac{19}{19+49} = 0.28$$

$$\text{F1} = 2 \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} = 0.97$$

Confusion matrix: Case3

Imagine you're building a system to detect emotions via audio. Among six classes of emotions, you need to detect sadness. The confusion matrix won't change much:



	Su	A	Sa	H	D	F	
Surprise							
Anger	TN		FP		TN		
Sadness	FN	FN	TP	FN	FN	FN	target
Happiness							
Disgust	TN		FP		TN		
Fear							

TN – True Negative, TP – True Positive
FN – False Negative, FP – False Positive

Confusion matrix: errors

H_0 – patient doesn't have COVID-19.

Type I error (first kind error; false positive) – your prediction/observation is positive, and it's false. Type I error involves rejecting a true null hypothesis – the patient has no COVID-19, but a PCR test falsely returned “positive.” Another example is a fire alarm going on, but there is no fire (“false alarm”).

Type II error (second kind error; false negative) – your prediction/observation is negative, and it's false. Type II error deals with failing to reject a false null hypothesis – the patient has COVID-19, but the PCR test fails to detect it. Another example: fire alarm, silent, whereas fire, is all around the building.

Usually, type II errors are far worse than type I errors, as the consequences of type II errors – not detecting the disease or fire is more harmful than setting up a false alarm.

Confusion matrix: metrics

Of course, besides the discussed metrics we can use basic accuracy:

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

or ROC-AUC: The Area Under the Curve (AUC) of the Receiver Operator Characteristic (ROC) probability curve. The ROC curve displays represent all confusion matrixes for a given model. The closer the AUC to 1 – the better the model.

We'll cover ROC-AUC deeper in one of the following lectures.

