Machine Learning

Model Evaluation

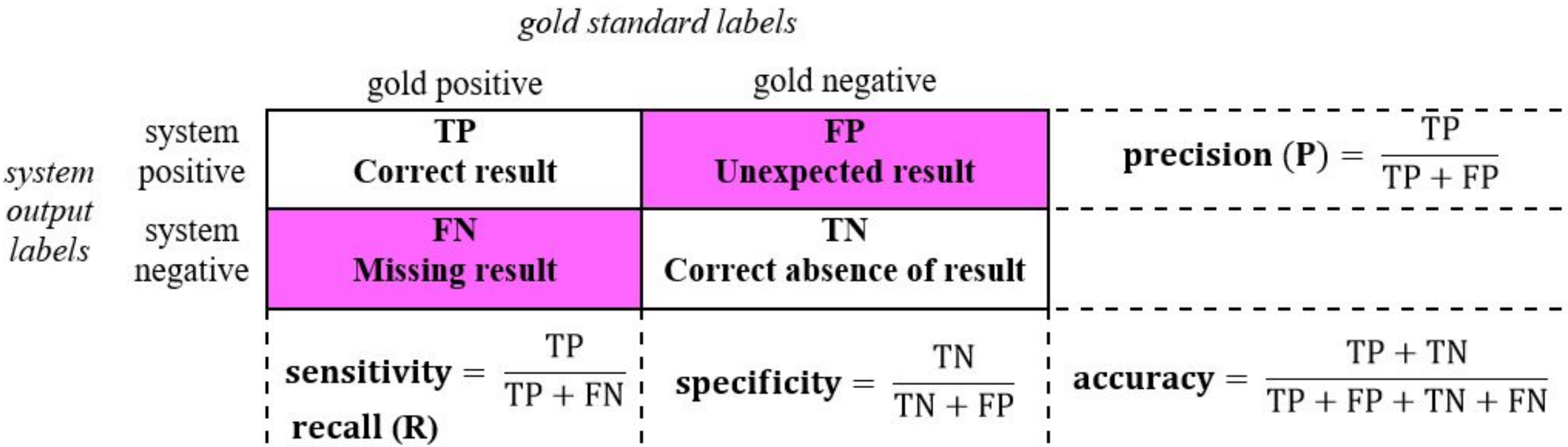
Tools

- There are a few main tools that are available to test a classification model's quality:
    - Confusion matrices (or truth tables)
    - Lift charts
    - ROC (receiver operator characteristic) curves
    - AUC (area under the curve)

Confusion Matrix / Contingency Table

		gold standard labels	
		gold positive	gold negative
system output labels	system positive	**TP** **Correct result**	**FP** **Unexpected result**
	system negative	**FN** **Missing result**	**TN** **Correct absence of result**

Evaluation Metrics



F-measure

- There are many ways to define a single metric that incorporates aspects of both precision and recall.
- The simplest of these combinations is the **F-measure** defined as:

$$F_\beta = \frac{(\beta^2+1)PR}{\beta^2 P+R}$$

- The β parameter differentially weights the importance of recall and precision, based perhaps on the needs of an application.

F-measure

- **F-measure** defined as:

$$F_\beta = \frac{(\beta^2+1)PR}{\beta^2 P+R}$$

- Values of $\beta > 1$ favor recall, while values of $\beta < 1$ favor precision.
- When $\beta = 1$, precision and recall are equally balanced; this is the most frequently used metric, and is called $F_\beta = 1$ or just F_1:

$$F_1 = \frac{2PR}{P+R}$$

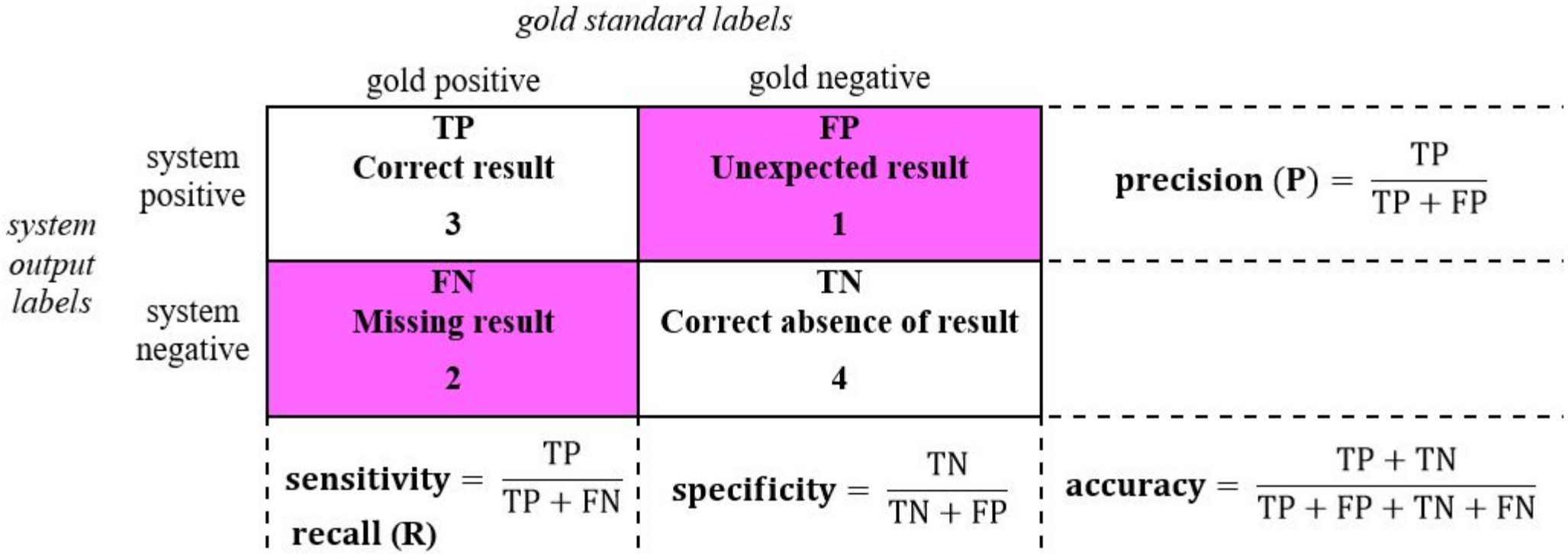
Model Evaluation

- Output of a classifier for email spam filtering.

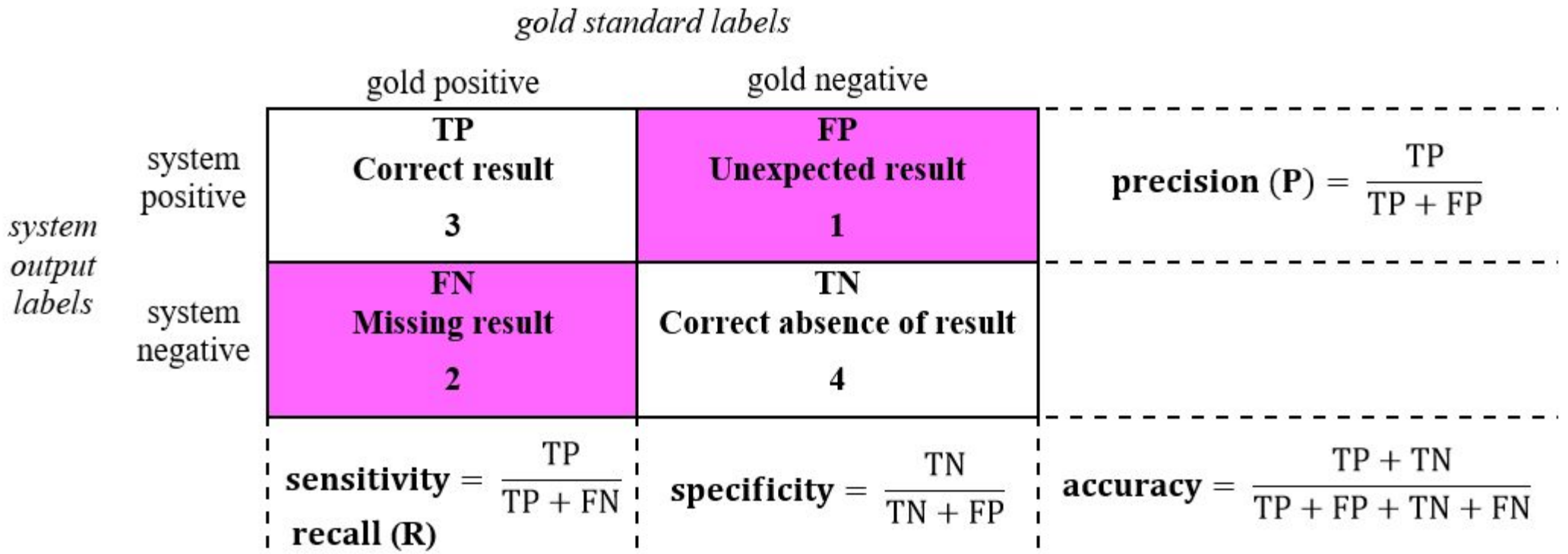
Sl.	Target	Prediction
1	Spam	Not Spam
2	Spam	Spam
3	Not Spam	Not Spam
4	Spam	Spam
5	Not Spam	Spam
6	Not Spam	Not Spam
7	Not Spam	Not Spam
8	Spam	Spam
9	Spam	Not Spam
10	Not Spam	Not Spam

Model Evaluation

- Calculate Sensitivity, Specificity, Precision, Recall, Accuracy, F1 score of this email spam filtering classifier.

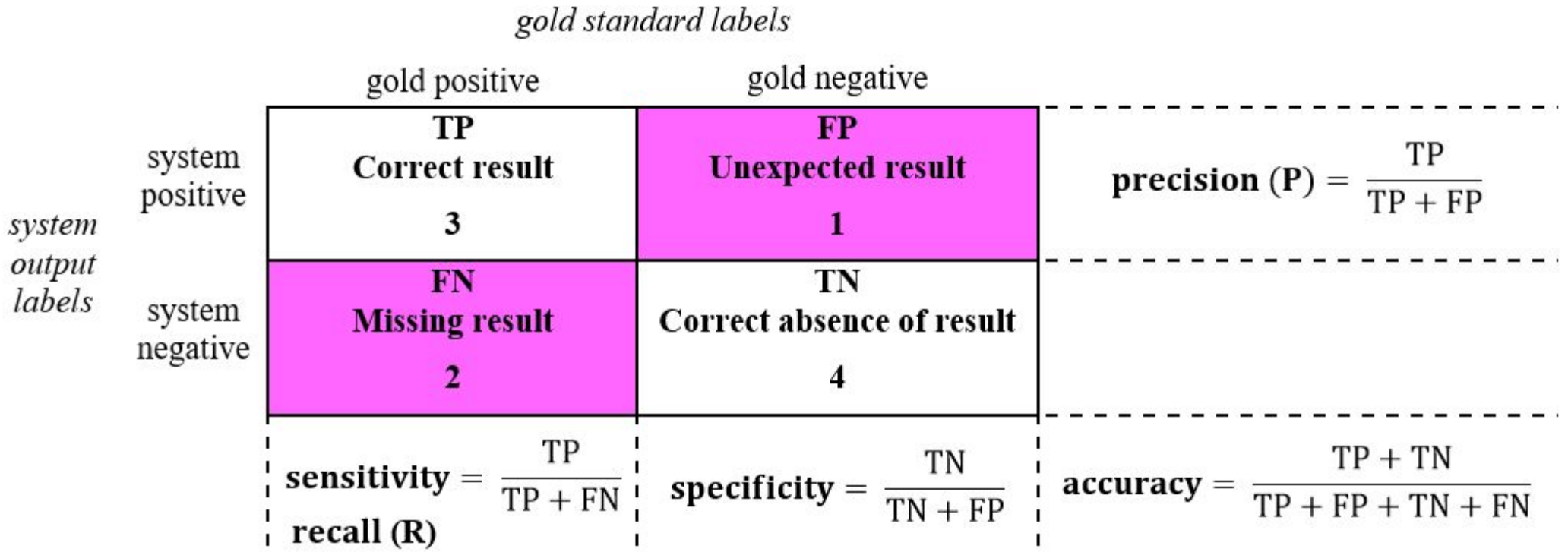


Sensitivity or Recall (R)



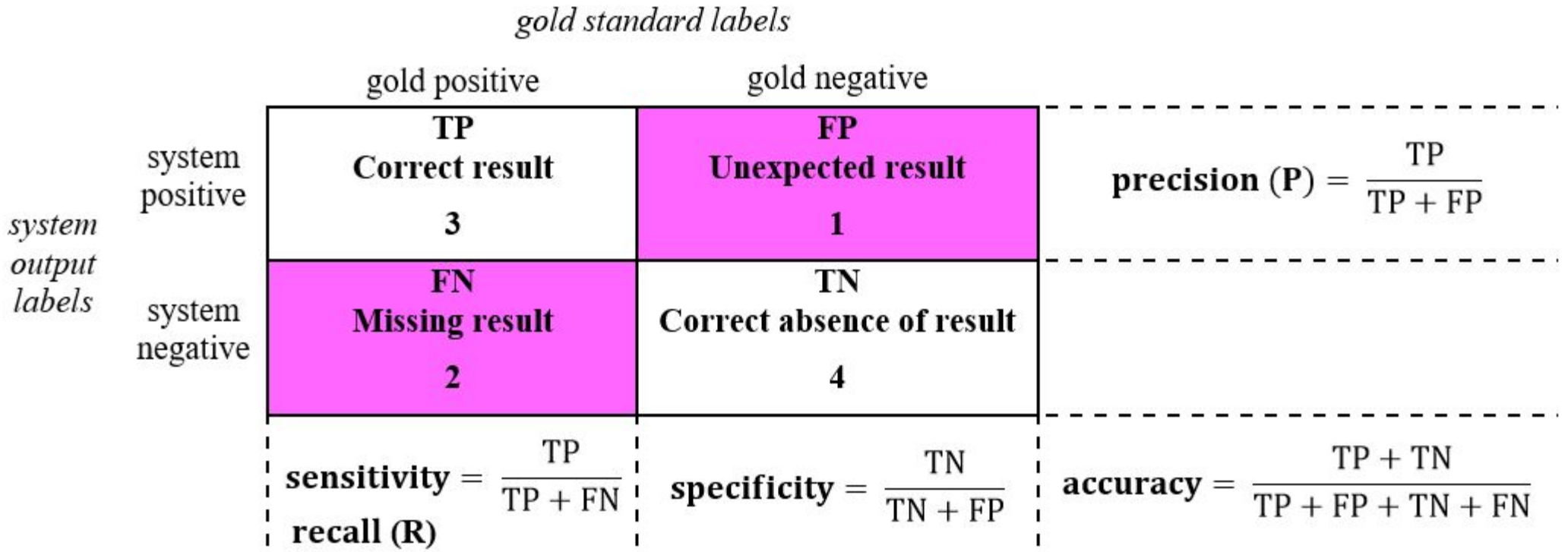
- Sensitivity or Recall (R) $= \frac{TP}{TP+FN} = \frac{3}{3+2} = \frac{3}{5} = 0.6$

Specificity



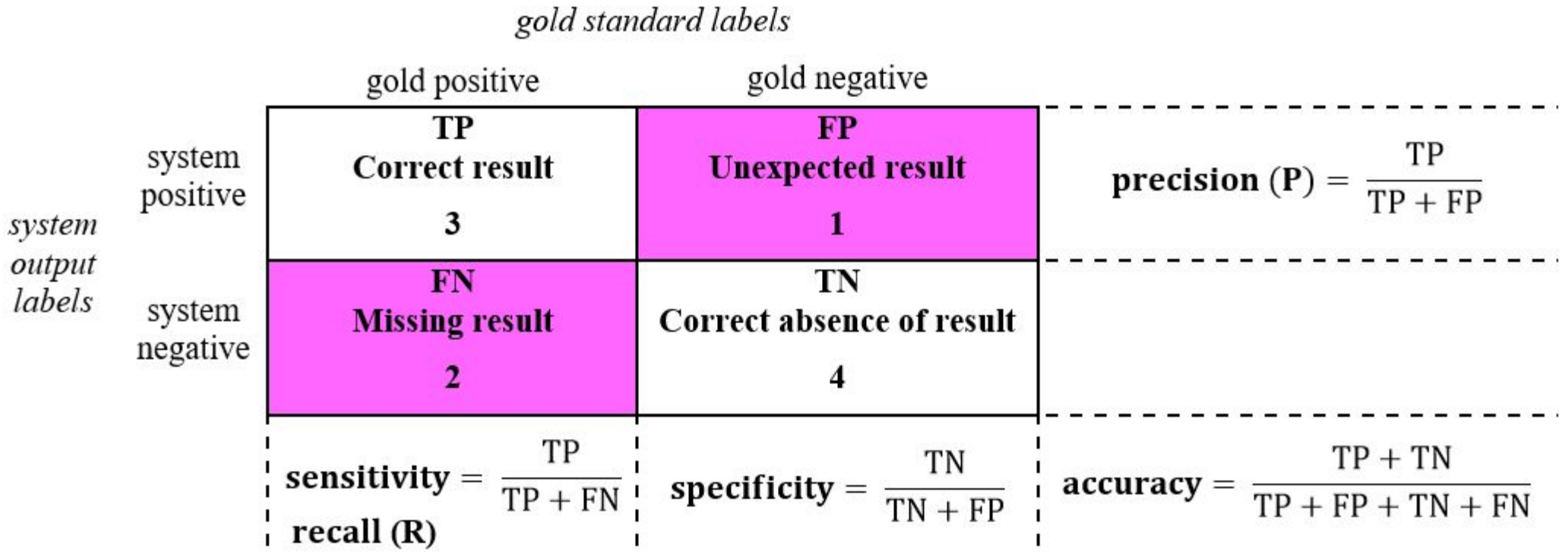
- Specificity $= \frac{TN}{TN+FP} = \frac{4}{4+1} = \frac{4}{5} = 0.8$

Precision



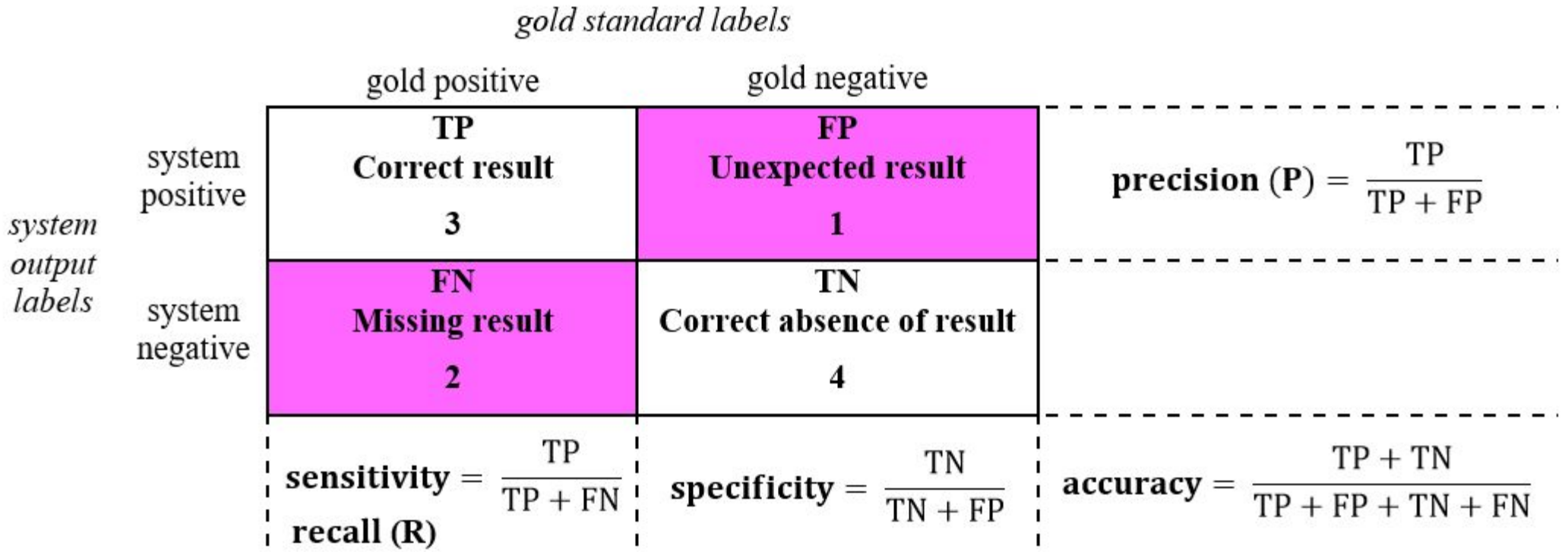
- Precision (P) $= \frac{TP}{TP+FP} = \frac{3}{3+1} = \frac{3}{4} = 0.75$

Accuracy



- $\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} = \frac{3+4}{3+1+4+2} = \frac{7}{10} = 0.7$

F-measure

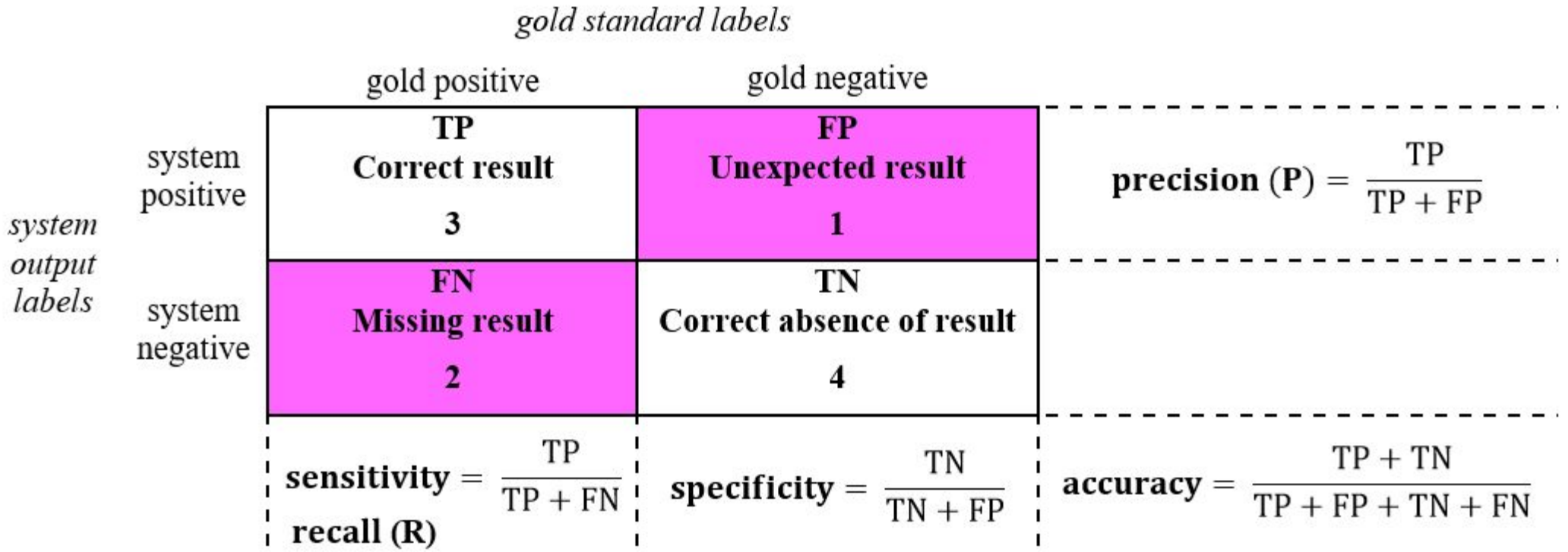


-
- F-measure for $\beta = 1$, $F_1 = \frac{2PR}{P+R} = \frac{2\times0.75\times0.6}{0.75+0.6} = \frac{0.9}{1.35} = 0.67$

ROC Curve

- A Receiver Operating Characteristic (ROC) curve is a graphical representation of a model's ability to distinguish between two classes (positive and negative) at different classification thresholds.
- It plots the True Positive Rate (sensitivity) against the False Positive Rate (1 - specificity).

F-measure



- True Positive Rate (TPR) = Sensitivity or Recall
- False Positive Rate (FPR) = 1 – Specificity = $\frac{FP}{FP+TN}$

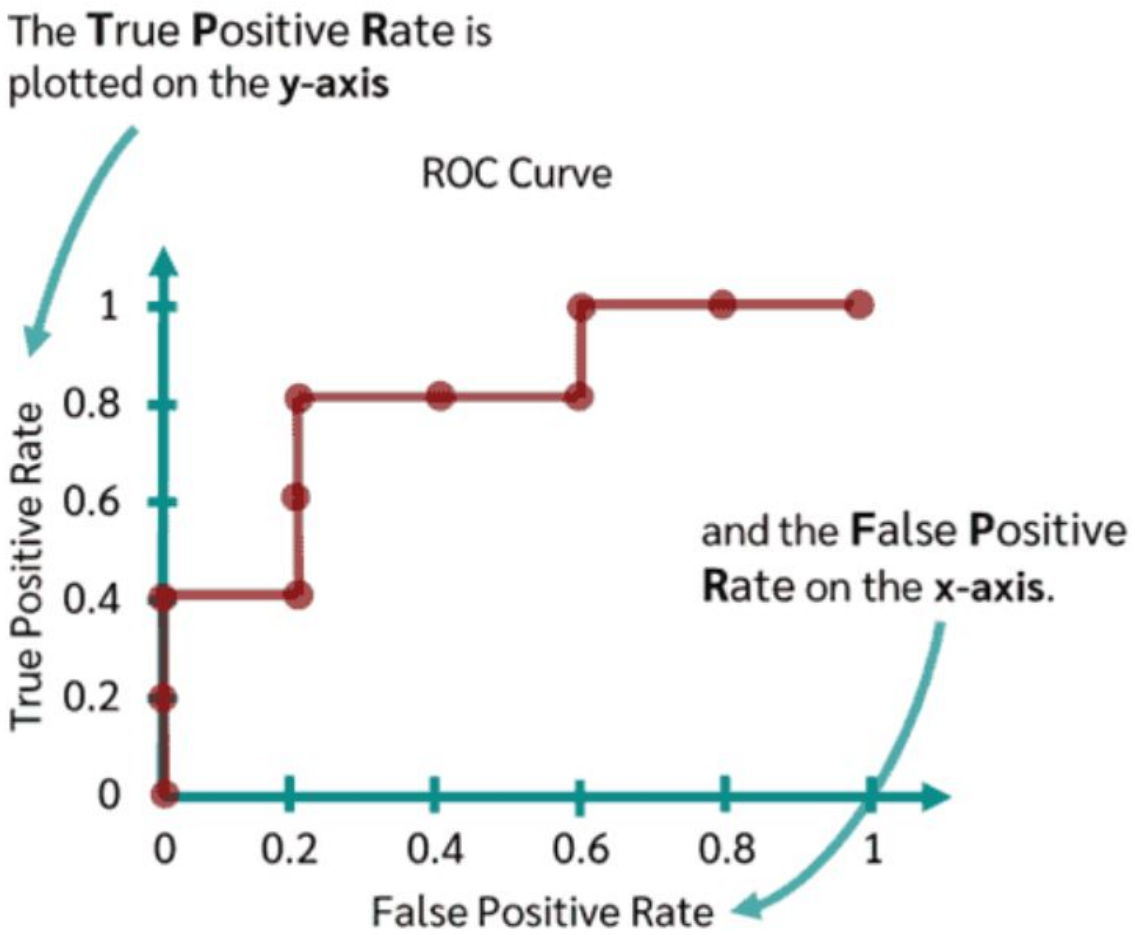
Example of a ROC Curve

- We would like to classify, based on a screening, whether a person has cancer or not.
- This classification is done with the help of a certain blood value, where high values indicate cancer.
- The question now is which value we choose as the classification threshold. So from which value do we predict a disease?
- For this, we obtain data from 10 people about how high the blood value is and whether or not the disease is present.

ROC Curve

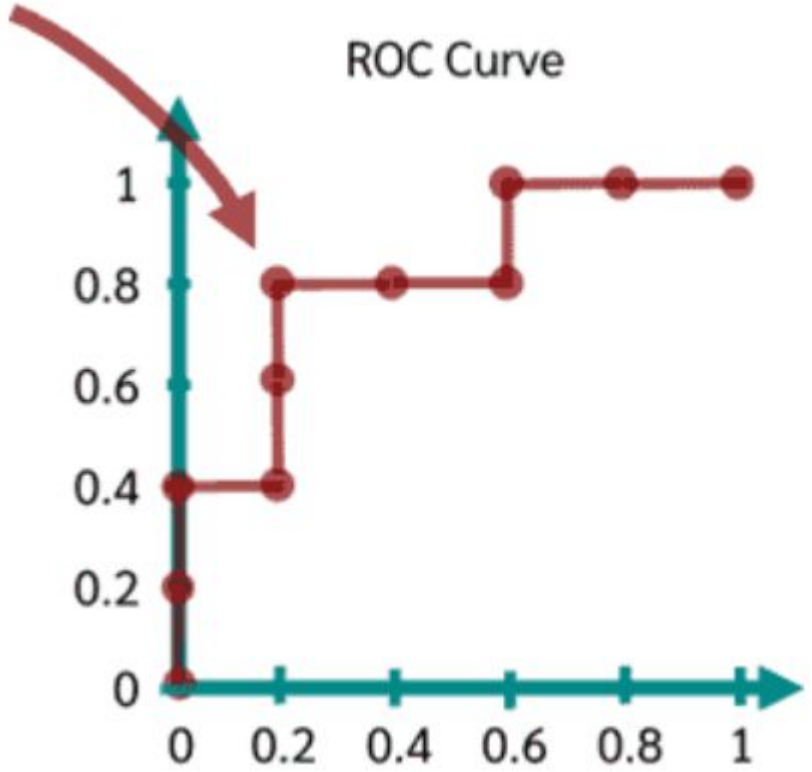
- We can now calculate for each threshold what the True Positive Rate and the False Positive Rate are.
- These two values are plotted on the ROC curve.
- The True Positive Rate is plotted on the *y-axis* and the False Positive Rate on the *x-axis*.

ROC Curve


The True Positive Rate is plotted on the y-axis
ROC Curve
True Positive Rate
1
0.8
0.6
0.4
0.2
0
0
0.2
0.4
0.6
0.8
1
False Positive Rate
and the False Positive Rate on the x-axis.

ROC Curve

- The curve visually illustrates the trade-off between correctly identifying positive cases and incorrectly identifying negative cases.
- At the marked point below, for example, 80% of the diseased people were correctly classified as "diseased" and 20% of the healthy people were incorrectly classified as "diseased".



ROC Curve

- Using the ROC curve, we can compare different classification methods. A classification model is better the higher the curve is.