AI in Medicine I
WiSe 2022/23

Technical University Munich
Department of Informatics

I31 - AI in Medicine and Healthcare
I32 - Computational Imaging in AI and Medicine

Theoretical exercise 7 **9. Jan 2023**

Risk Scores and Stratification

The solutions will be discussed in the tutorial session

12. Jan 2023, 4-6 p.m. in lecture hall 5901.EG.051

For questions regarding this exercise sheet, please contact: `haifa.beji@tum.de`
For general questions, please contact: `course.aim-lab@med.tum.de`

1 General Questions

1. What is a risk score? And what does stratification mean?

2. Define the *Relative Risk* and *Odds Ratio*? How do they compare to each other?

3. You have been entrusted with a large database of patients and want to quantify the risk of the patients developing skin cancer within the next 25 years into low, moderate or high risk. What kind of machine learning problem would you formulate this as?

   ◯ clustering
   ◯ segmentation
   ◯ classification
   ◯ regression

2 Diagnostic Tests Assessment

Your collaborators are investigating a marker as a potential screening test for ovarian cancer. They collected a dataset of subjects and noted how the results of the tests compare to the actual cancer diagnoses, resulting in the table below.

	Disease	No Disease
Test Positive	100	200
Test Negative	10	1000

Compute the following:

- precision / positive predictive value
- recall / sensitivity
- negative predictive value
- specificity / true negative rate
- accuracy
- Dice score: $\dfrac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$
- F1 score: $\dfrac{2}{\dfrac{1}{\text{precision}} + \dfrac{1}{\text{recall}}}$

In another experiment the threshold at which the test is considered positive was altered, resulting in more positive test results, summarized in the following table:

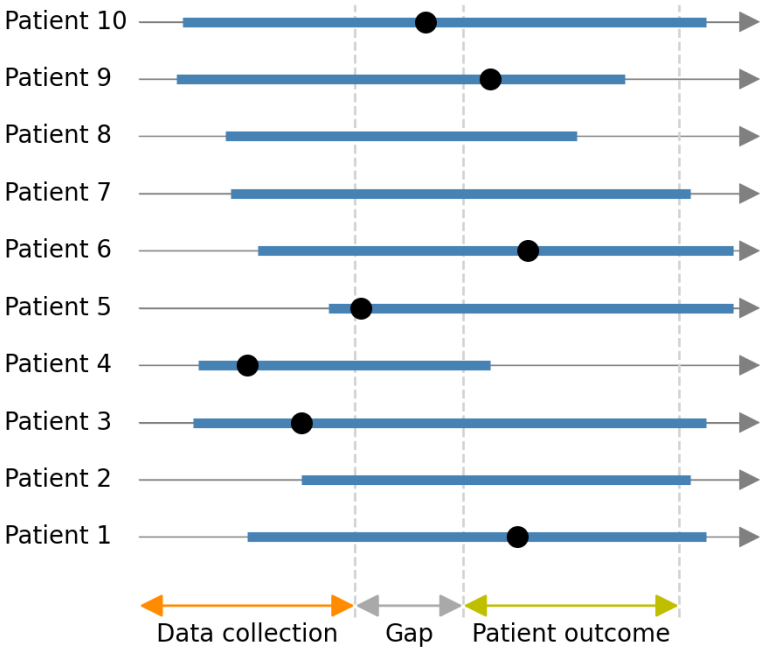
	Disease	No Disease
Test Positive	105	400
Test Negative	5	800

- How do the two thresholds compare in terms of precision and recall?
- Which threshold is better?

3 Dataset Construction

We want to compute diabetes risk for patients at the local hospital. A data collection period of 3 years, followed by a gap period of 1 in year and an evaluation period of 3 years is agreed upon. We visualize a sample of the patients aligned by absolute time, where a black dot denotes disease onset for a given patient. Indicate which patients should be included and which should be excluded from the dataset by assigning them to the following categories:

1. Include, negative sample
2. Include, positive sample
3. Exclude, disease onset before evaluation period
4. Exclude, right censoring
5. Exclude, left censoring



4 Receiver Operator Characteristic (ROC) curve

Below are all of the predicted probabilities of a binary classifier before binarization, sorted according to the probability scores $p_i \in [0, 1]$. The ground truth label for each point is indicated by its color (red or green), but the predicted label is only to be determined after choosing a cutoff threshold T (eg. predicted label$(i) = 1 \iff p_i \geq T$).

For each of the thresholds t_1, t_2, t_3 and t_4 shown in the top figure, compute the corresponding point on the ROC curve (FPR, TPR) and add it to the ROC plot (bottom figure).