Performance Indices for Classification

J.-S. Roger Jang (張智星)
MIR Lab, CSIE Dept.
National Taiwan University
jang@mirlab.org, http://mirlab.org/jang

2023/11/29

Outline

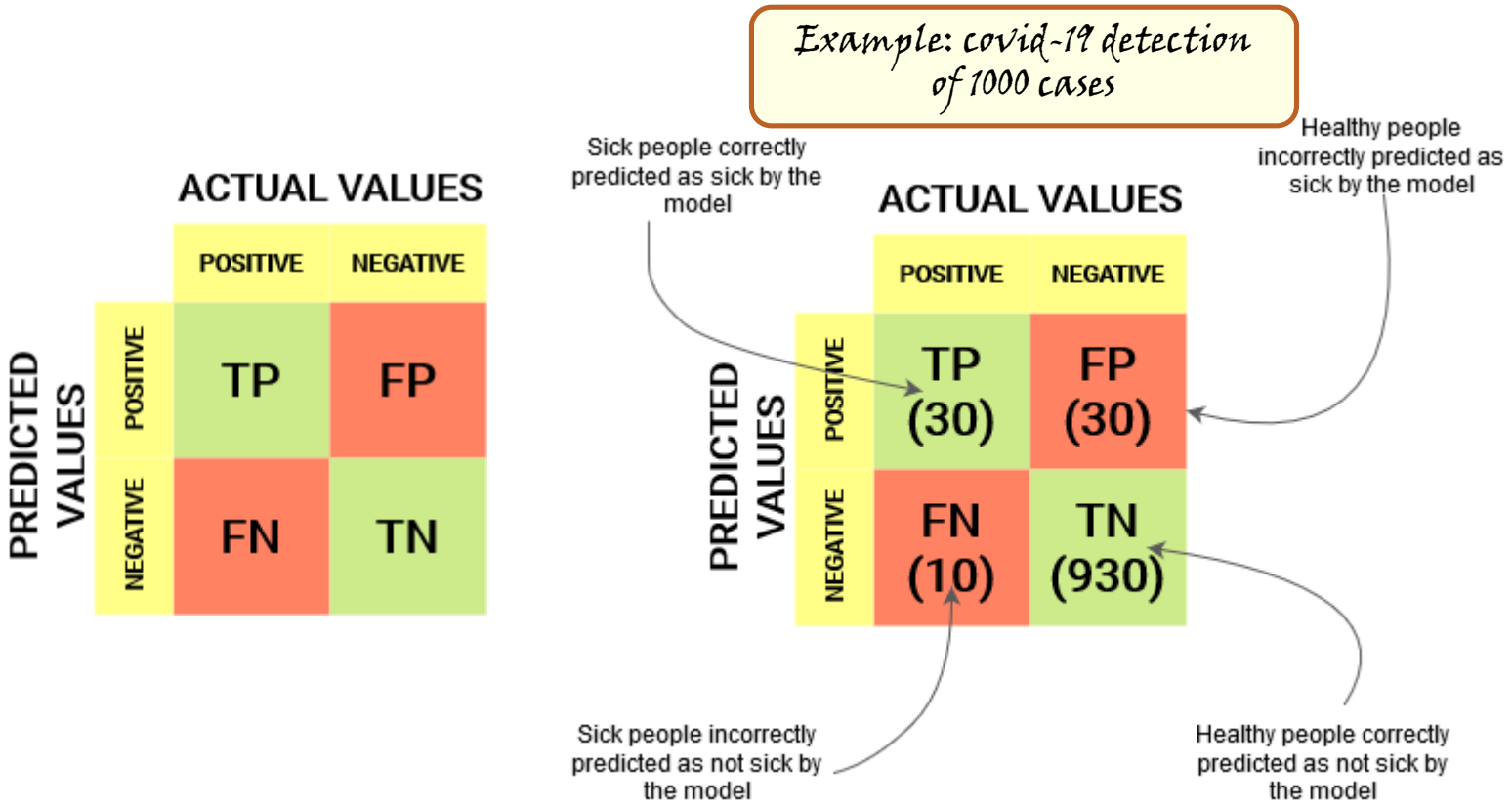
- Confusion matrices
  - Performance indices based on confusion matrix
- ROC, DET and AUROC
  - Cost-sensitive classification
- PRC and AUPRC
  - Lift charts
- Exercises

Intro. to Performance Indices (PIs)

- Performance index
  - AKA performance metrics, figure of merit, etc.
  - To evaluate the performance of a ML model
- Different situations require different PIs
  - Purposes
    - Classification
    - Sequence decoding
    - Regression
    - Ranking
  - Situations
    - Imbalanced dataset for classification
    - Binary or multi-class classification problems

Confusion Matrix for Binary Classification

- Confusion matrix for binary classification
  - Can be extended to multi-class classification

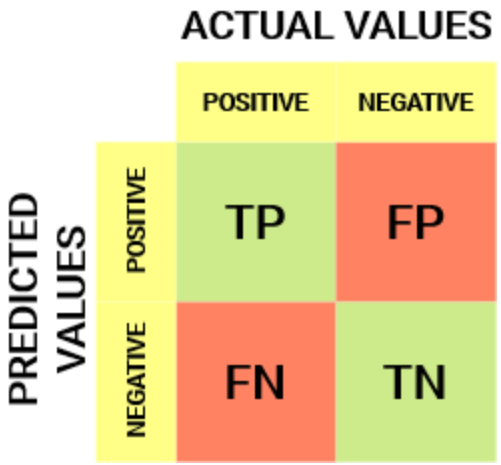


https://www.analyticsvidhya.com/blog/2020/04/confusion-matrix-machine-learning

PIs Derived from Confusion Matrices

Quiz!

- Numerous PIs can be derived from a confusion matrix!
  - Accuracy, recognition rate
    - $\text{Accuracy} = \frac{TP+TN}{All}$
  - True positive rate, sensitivity, hit rate, recall
    - $\text{TPR} = \frac{TP}{P} = \frac{TP}{TP+FN} = 1 - FNR$
  - True negative rate, specificity, selectivity
    - $\text{TNR} = \frac{TN}{N} = \frac{TN}{FP+TN} = 1 - FPR$
  - False positive rate, miss rate
    - $\text{FPR} = \frac{FP}{N} = 1 - TNR$
  - False negative rate
    - $\text{FNR} = \frac{FN}{P} = 1 - TPR$
  - Precision
    - $\text{Precision} = \frac{TP}{TP+FP}$



https://en.wikipedia.org/wiki/Confusion_matrix

Precision and Recall

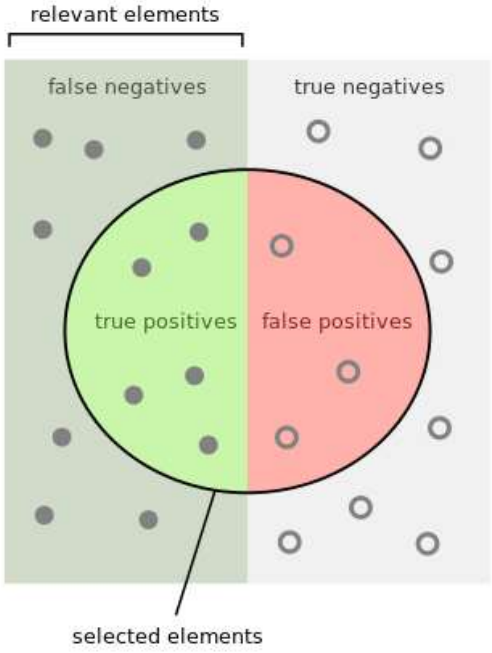
Quiz!

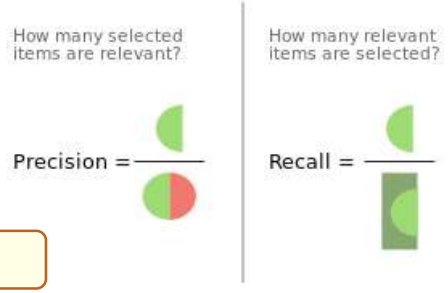
- From the viewpoint of information retrieval
  - Precision: $\mathrm{p} = \frac{TP}{TP+FP}$
  - Recall: $\mathrm{r} = \frac{TP}{TP+FN}$
  - F-measure: $\mathrm{f} = \frac{2pr}{p+r} = \frac{TP}{(2TP+FP+FN)/2}$

Harmonic mean of p and r

Aka F1-score

- Usually
  - p ↗ ➔ r ↘
  - P ↘ ➔ r ↗





https://en.wikipedia.org/wiki/Precision_and_recall

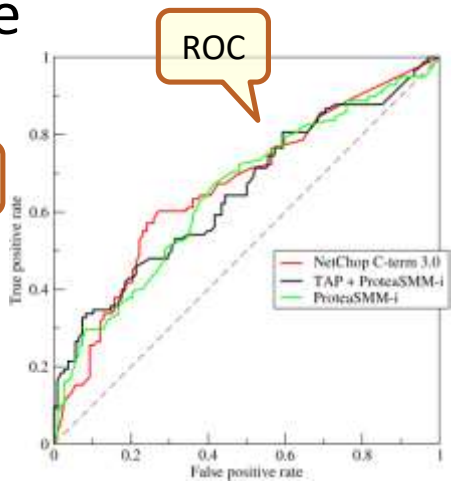
ROC and DET

Quiz!

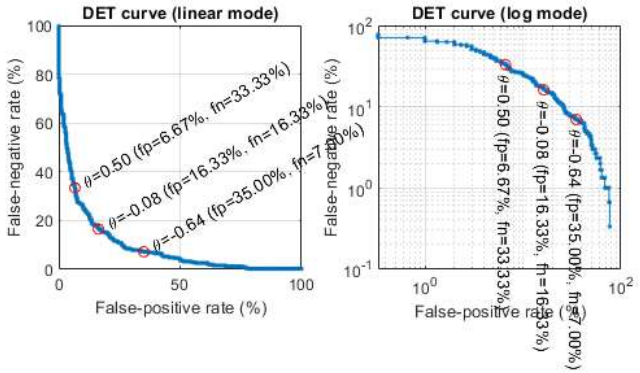
- ROC (Receiver Operating Characteristics) curve
  - TPR vs. FPR when threshold is varying
  - Endpoints at (0, 0) & (1, 1)
- DET (Detection Error Tradeoff) curve
  - FNR vs. FPR when threshold is varying
  - Endpoints at (0, 1) & (1, 0)

Similarity, likelihood, etc

ROC



DET

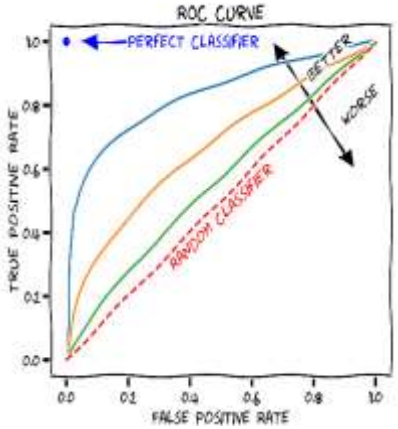


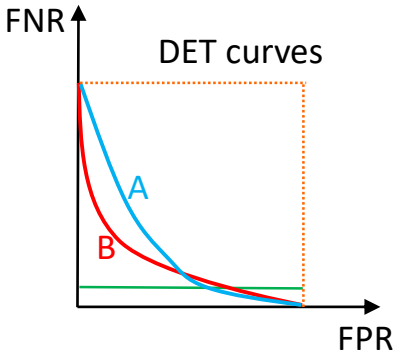
https://en.wikipedia.org/wiki/Receiver_operating_characteristic

AUROC

Quiz!

- AUROC: Area under ROC
  - 0.5 ≦ AUROC ≦ 1.0
    - 0.5 ➔ Random guess
    - 1.0 ➔ Perfect classification
  - Suitable PI for imbalance datasets
  - Determine the operating point based on cost
    - For covid-19 detection, which classifier is better, A or B? At what operating points?

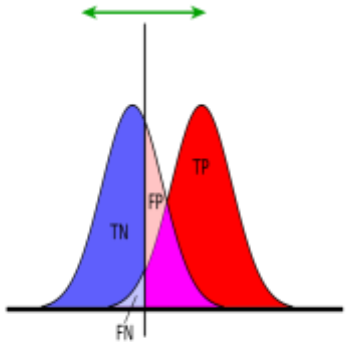




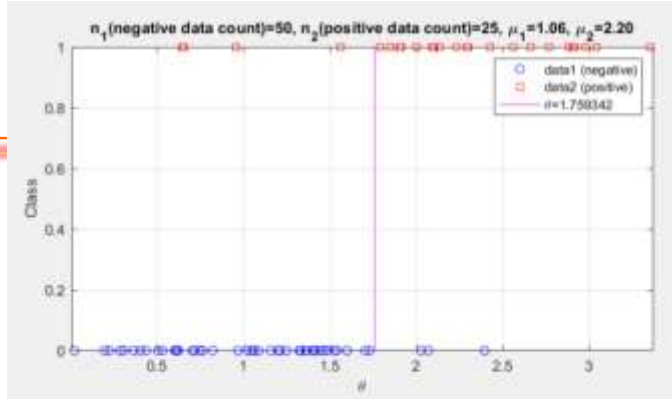
https://en.wikipedia.org/wiki/Receiver_operating_characteristic

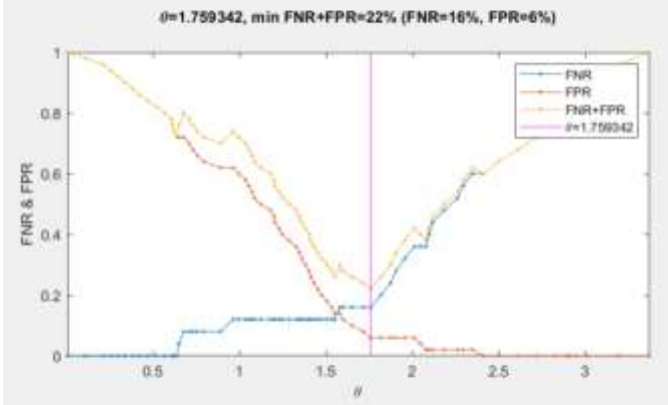
Two Ways to Plot ROC

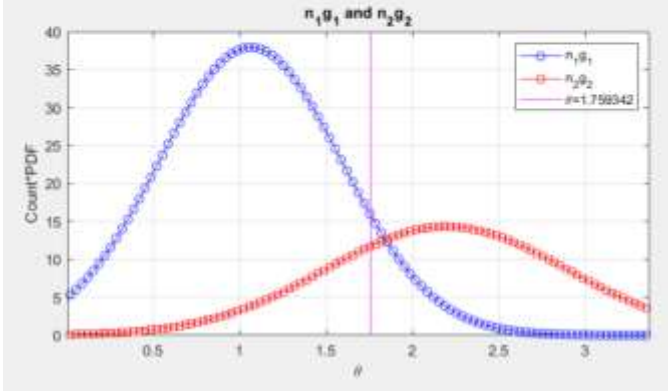
- Sample-point based
  - Less smooth
  - More realistic
- Distribution based
  - More smooth
  - If the threshold is at the intersection
    - FNR+FPR is minimized
    - TPR+TNR is maximized



TP	FP
FN	TN



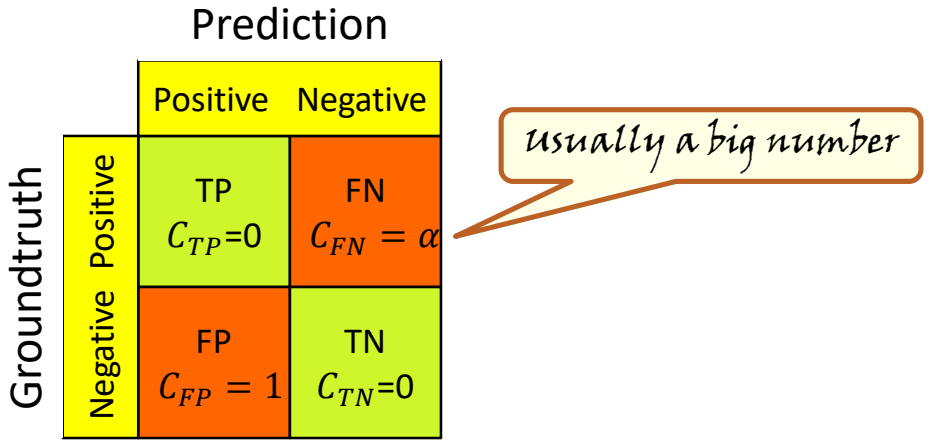




Cost-sensitive Classification

Quiz!

- Cost matrix

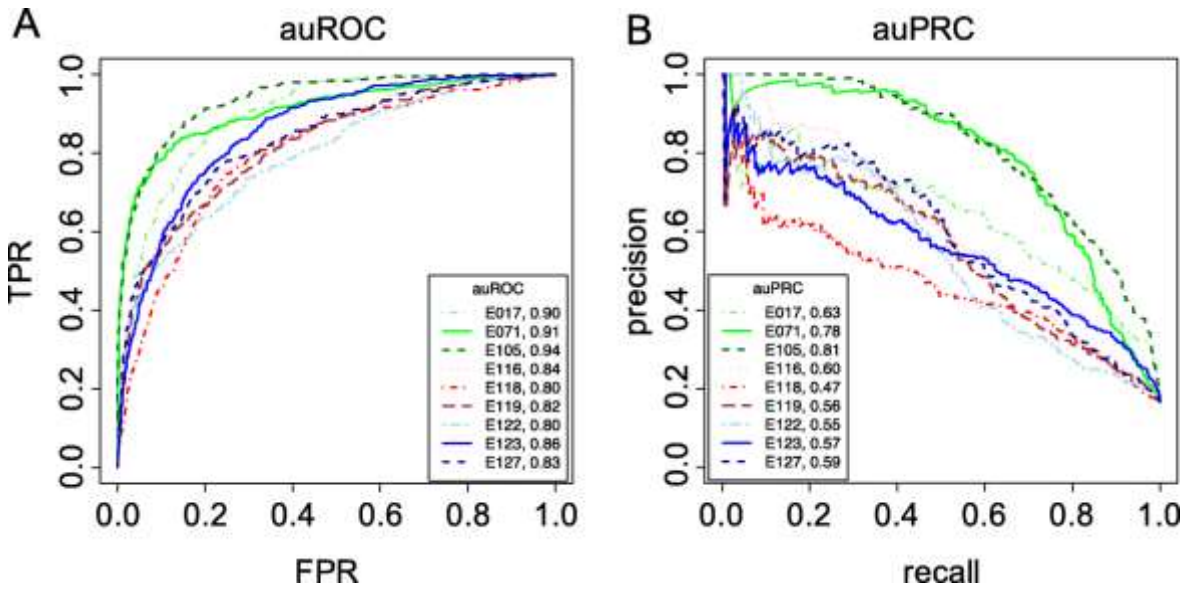


- Objective function
  - $\mathrm{f}(\theta) = C_{TP} * \mathrm{TP}(\theta) + C_{FN} * \mathrm{FN}(\theta) + C_{FP} * \mathrm{FP}(\theta) + C_{TN} * \mathrm{TN}(\theta)$
    $= \alpha * \mathrm{FN}(\theta) + \mathrm{FP}(\theta)$
  - $\hat{\theta} = \arg\min_{\theta} f(\theta)$

Precision-Recall Curve (PRC)

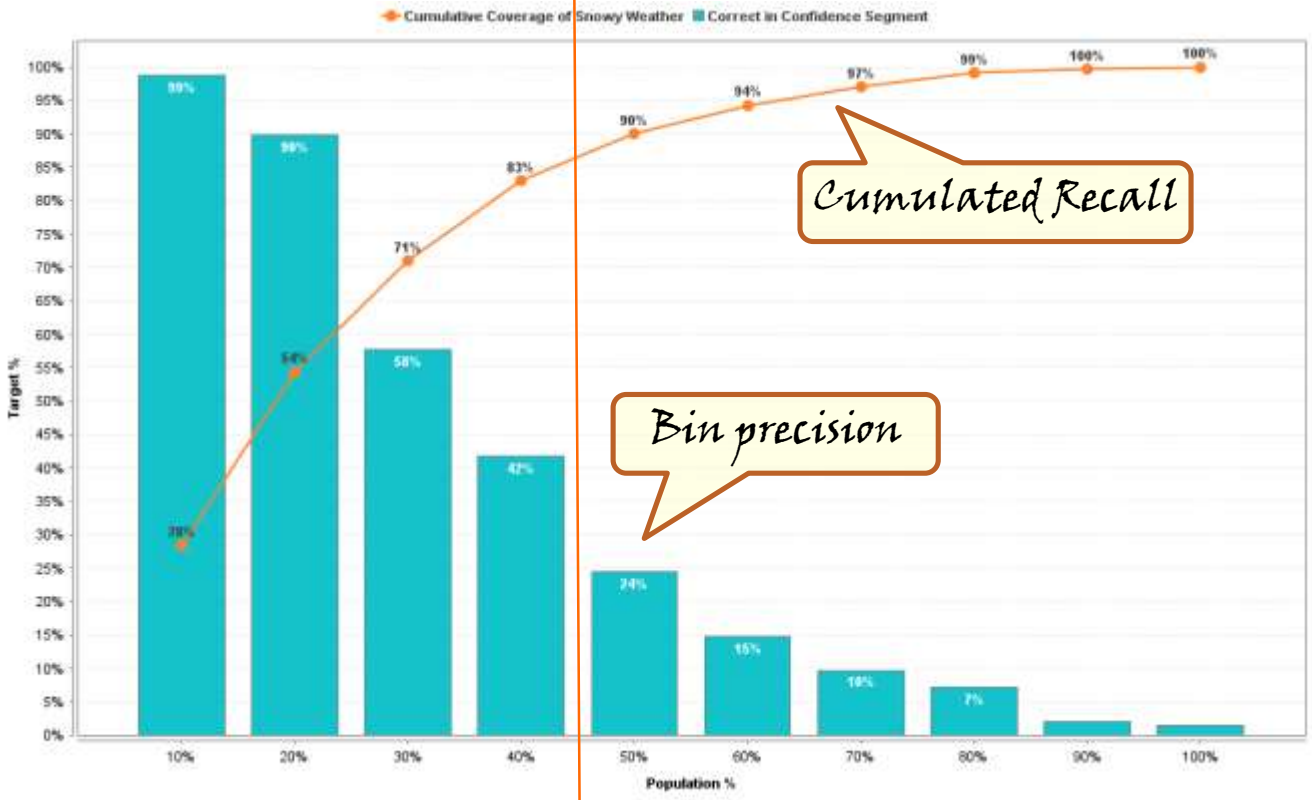
- PRC
  - Plot of precision vs. recall
- AUPRC (area under PRC)
  - Usually not smooth
  - Where are the endpoints?

Quiz!



Lift Charts

- Good for convincing people to appreciate power of ML
- Example:



← 對40%人發送廣告，就可以抓到83%想出國的人！

Accuracy of Sequence Decoding

- Accuracy of ASR (automatic speech recognition) for English
  - WER: word error rate

$$WER = \frac{D + I + S}{N} \Rightarrow Accuracy = 1 - WER$$

Length of groundtruth

Could be negative!

  - Example:

Human-labeled Transcript: How are you today John
Speech Recognition Result: How you a today Jones

D I S

$$WER = \frac{1 + 1 + 1}{5} = 60\% \Rightarrow Accuracy = 40\%$$

- Facts
  - Accuracy could be negative! ➔ Counter intuitive!
  - You need to use edit distance (based on DP) to find D, I, and S. ➔ Computation intensive!

Error Rate of Sequence Decoding

Quiz!

- Error rates of ASR for Chinese and mixed code
  - Chinese
    - Groundtruth: 先填表格 再打疫苗
    - ASR result: 鮮甜表格 再打疫苗
    - CER= $\frac{2}{8}$= 25.0%
    - WER= $\frac{1+1}{6}$=33.3%
  - Chinese and English
    - Groundtruth: 我有點喜歡iPhone
    - ASR result: 我優點喜歡哀鳳
    - CER = $\frac{7}{11}$ = 63.6%
    - WER = $\frac{4}{5}$ = 80.0%
    - MER= $\frac{3}{6}$ = 55.7%

WER: Word error rate
(Chinese & English: word)
CER: Character error rate
(Chinese & English: character)
MER: mixed error rate
(Chinese: character, English: word)

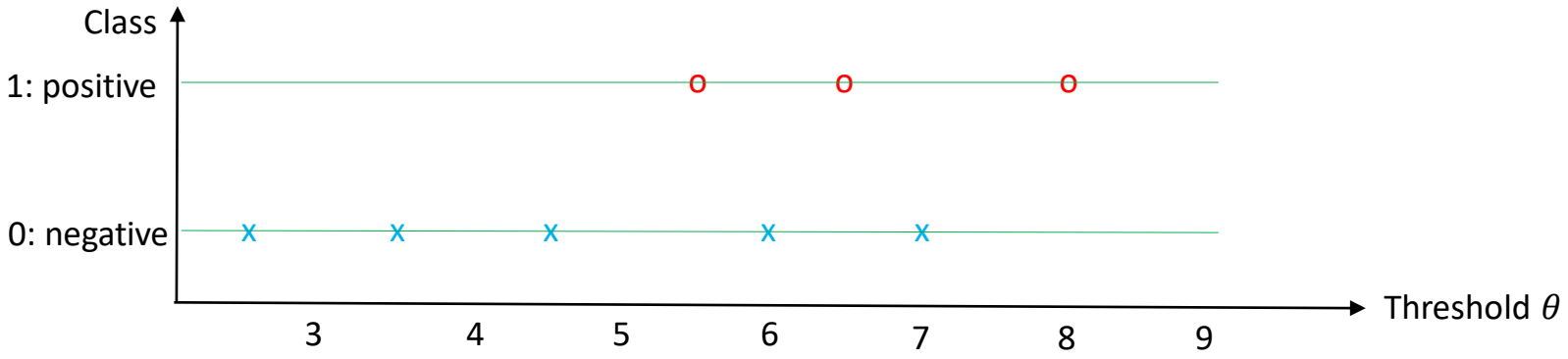
Exercise: PIs Based on a Confusion Matrix

- Derive various PIs based on the confusion matrix obtained from a detection method for covid-19:

		Prediction: Positive	Prediction: Negative
Groundtruth	Positive	TP = 30	FN = 10
	Negative	FP = 20	TN = 940

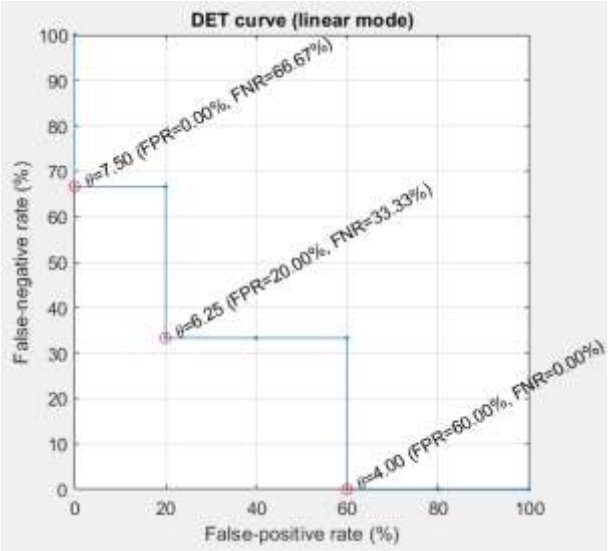
Exercise: ROC/DET/PRC

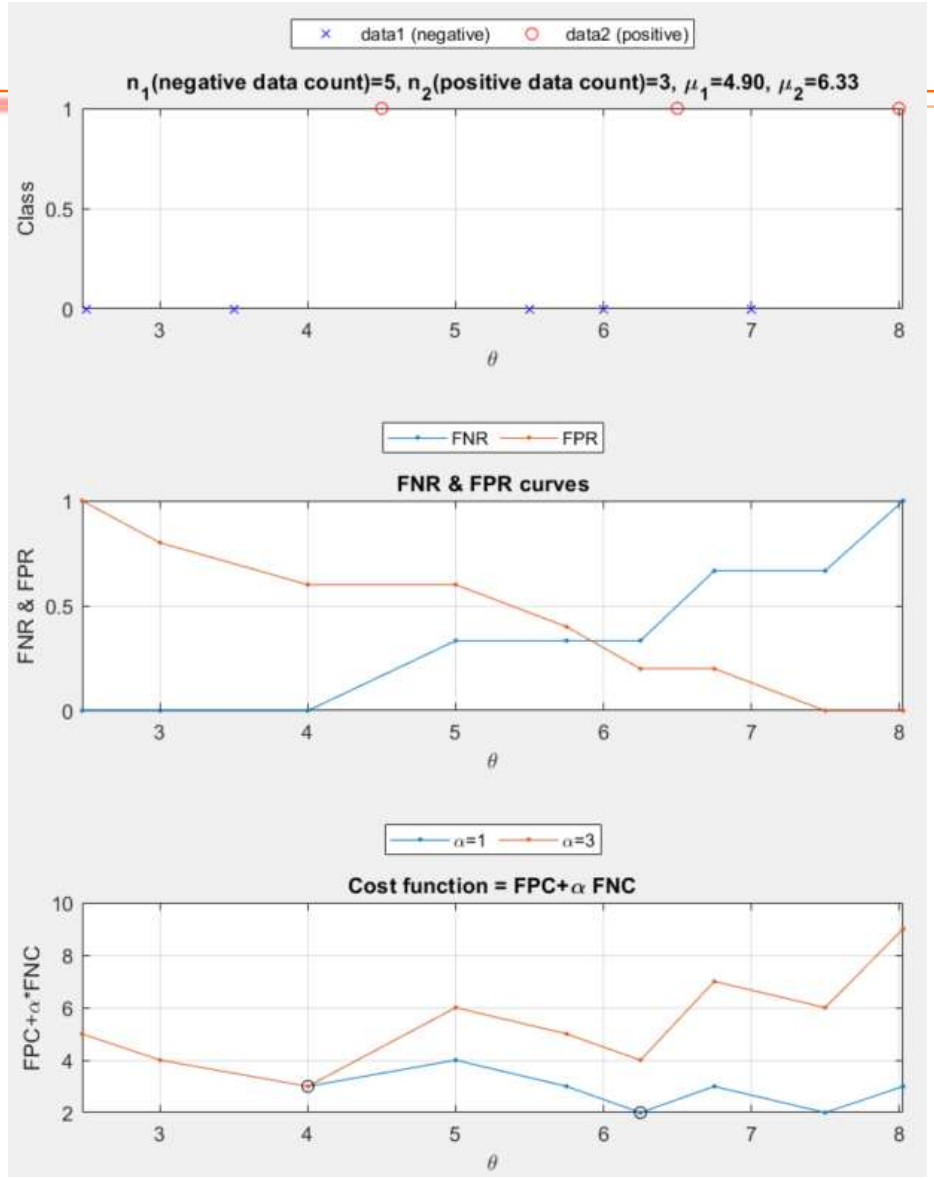
- A plot of sample data vs. thresholds is shown next.
  - Plot FNR/FPR vs. thresholds, ROC, DET, PRC.
  - Plot $\mathrm{f}(\theta) = \alpha * \mathrm{FN}(\theta) + \mathrm{FP}(\theta)$ when α=1 and 3, respectively. What is the corresponding $\arg\min_{\theta} f(\theta)$?



Exercise: DET/ROC

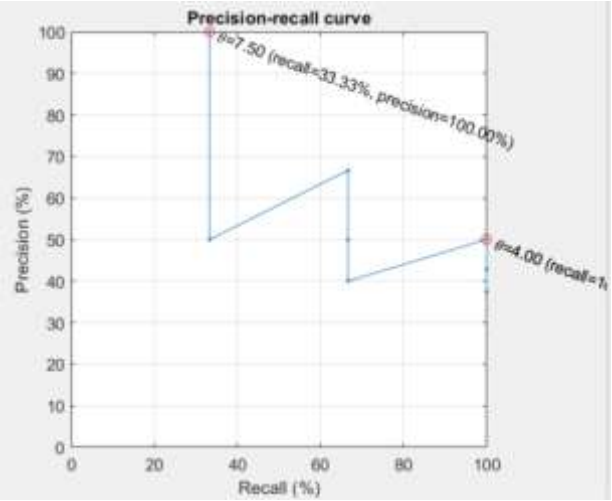
- Another example
  - FNR & FPR curves
  - DET plots
  - Cost function

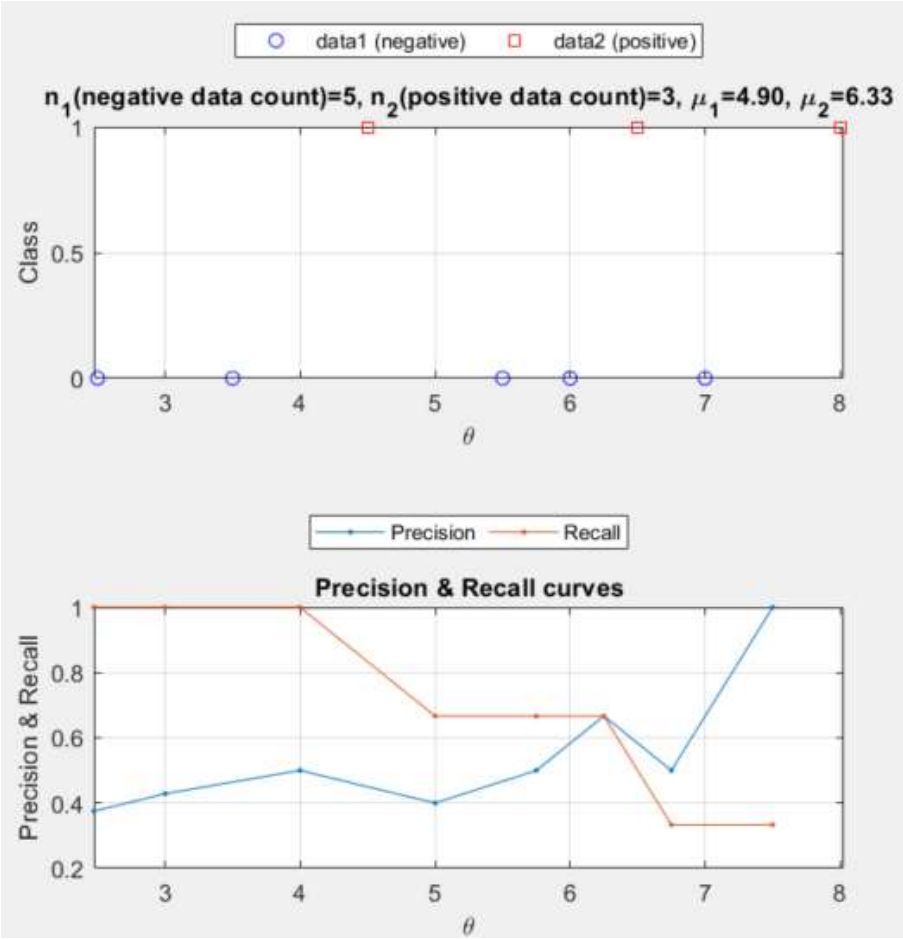




Exercise: PRC

- Another example
  - Precision & recall curves
  - PRC plots



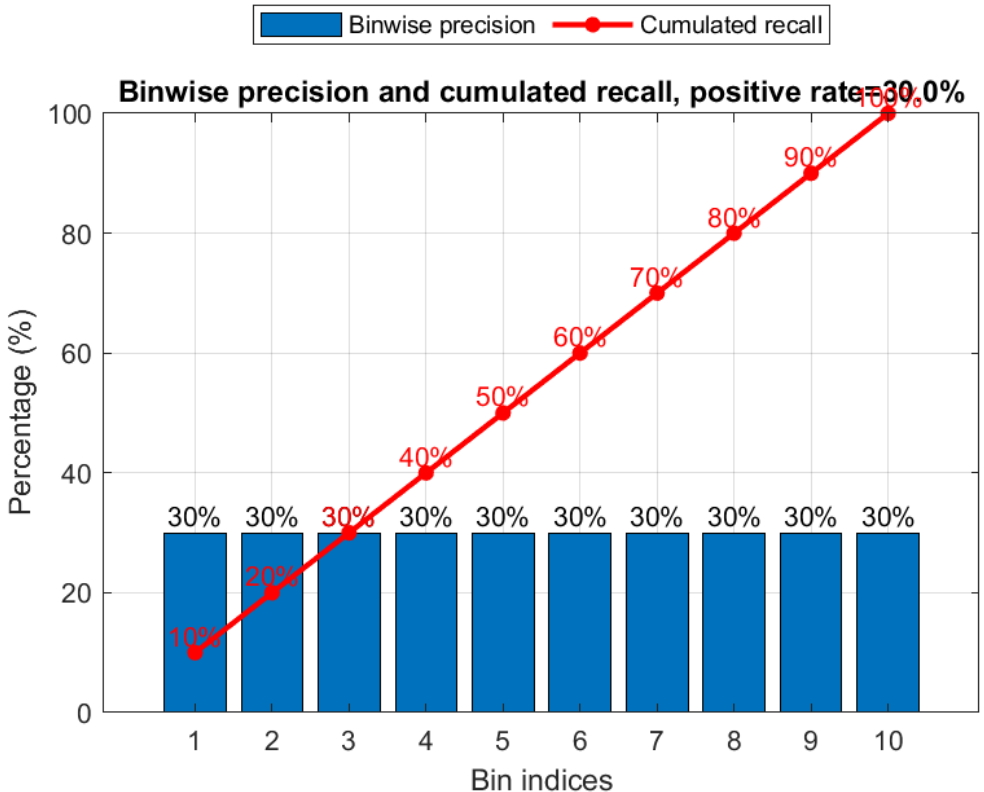


Exercise: DET/PRC

- Increasing? Decreasing? Or none of the above?
  - FNR vs. threshold
  - FPR vs. threshold
  - DET (FNT vs. FPR)
  - Precision vs. threshold
  - Recall vs. threshold
  - PRC (precision vs. recall)
- Endpoints?
  - DET
  - PRC

Exercise: Lift Chart

- Plot the ideal lift chart of a random-guess classifier, with no. of bins equal to 10, and N:P=7:3 where N and P are numbers of negative and positive cases, respectively.



Exercise: Yet Another Lift Chart

- Plot the lift chart of a perfect classifier when P=335 and N=665, with no. of bins equal to 10.