香港城市大學
City University of Hong Kong
專業 創新 胸懷全球
Professional·Creative
For The World

Classification:
Different Evaluation Metrics

CS5483 Data Warehousing and Data Mining

Medical diagnosis
From patient's perspective

- A test for COVID19 has accuracy 90%, i.e.,
$$\Pr(\hat{Y} = Y) = 0.9$$
  - Y: Indicator of infection.
  - $\hat{Y}$: Diagnosis of infection.
- Suppose a person is diagnosed to have the virus, i.e., $\hat{Y} = 1$.
  - Is it likely ($> 50\%$ chance) that the person has the virus? Y/N
  - Is the likelihood 90%? Y/N

Confusion matrix for binary classification

D	Predicted +ve, $\hat{Y} = 1$	Predicted -ve, $\hat{Y} = 0$
Actual +ve, $Y = 1$	TP	FN
Actual –ve, $Y = 0$	FP	TN

- **TP** (**True +ve**): number of +ve tuples classified as +ve.
- **TN** (**True -ve**): number of -ve tuples classified as -ve.
- **FP** (**False +ve**): number of -ve tuples classified as +ve.
  (**F_____ a_____ / Type I error**)
- **FN** (**False -ve**): number of +ve tuples classified as -ve.
  (**M_____ d_____ / Type II error**)

Accuracy vs Precision

$D\ (n)$	Predicted +ve ($\widehat{\mathrm{P}}$)	Predicted -ve
Actual +ve	TP	FN
Actual -ve	FP	TN

- **Accuracy** is $\frac{\mathrm{TP+TN}}{n}$ where $n = \mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}$.
- **Precision** is $\frac{\mathrm{TP}}{\widehat{\mathrm{P}}}$ where $\widehat{\mathrm{P}} = \mathrm{TP} + \mathrm{FP}$.

  P_____ p _____ v _____ (PPV)
- Is it possible that accuracy is high but precision is low?

Example

D (100)	Predicted +ve (10)	Predicted -ve (90)
Actual +ve	5	5
Actual -ve	5	85

- Accuracy is _____%.
- Precision is _____%.
- When is accuracy>precision in general?

Negative predictive value (NPV)

D (n)	Predicted +ve ($\widehat{\mathrm{P}}$)	Predicted -ve ($\widehat{\mathrm{N}}$)
Actual +ve	TP	FN
Actual -ve	FP	TN

- **NPV** is $\frac{\mathrm{TN}}{\widehat{\mathrm{N}}}$ where $\widehat{\mathrm{N}} = \mathrm{TN} + \mathrm{FN} = n - \widehat{\mathrm{P}}$.
- Accuracy is $\frac{\mathrm{TP}+\mathrm{TN}}{n} = \frac{\widehat{\mathrm{P}}\cdot\mathrm{PPV}+\widehat{\mathrm{N}}\cdot\mathrm{NPV}}{n} = \frac{\widehat{\mathrm{P}}}{n}\mathrm{PPV}+\frac{\widehat{\mathrm{N}}}{n}\mathrm{NPV}$.
  - Accuracy>precision iff NPV>/=/<PPV.
  - Accuracy=precision iff _____

Example

D (100)	Predicted +ve (10)	Predicted -ve (90)
Actual +ve	5	5
Actual -ve	5	85

- Accuracy is _____%.
- Precision is _____%.
- NPV is _____%.

Medical diagnosis
From Government's perspective

- Suppose the government wants to eradicate COVID19 as it is highly contagious.
- If a test is 90% accurate, can the government identify >50% of infected people? Y/N

Recall

$D\ (n)$	Predicted +ve ($\widehat{\mathrm{P}}$)	Predicted -ve ($\widehat{\mathrm{N}}$)
Actual +ve (P)	TP	FN
Actual -ve	FP	TN

- **Recall** is $\frac{\mathrm{TP}}{\mathrm{P}}$ where $\mathrm{P} = \mathrm{TP} + \mathrm{FN}$.
  S_____ or **True positive rate (TPR)**

Example

D (100)	Predicted +ve (10)	Predicted -ve (90)
Actual +ve (10)	5	5
Actual -ve	5	85

- Accuracy is _____%.
- Precision is _____%.
- NPV is _____%.
- Recall is _____%.
- When is accuracy>recall?

Specificity

$D\ (n)$	Predicted +ve ($\widehat{\mathrm{P}}$)	Predicted -ve ($\widehat{\mathrm{N}}$)
Actual +ve (P)	TP	FN
Actual -ve (N)	FP	TN

- **Specificity** is $\frac{\mathrm{TN}}{\mathrm{N}}$ where $\mathrm{N} = \mathrm{TN} + \mathrm{FP}$.
  True negative rate (TNR)
- Accuracy is $\frac{\mathrm{TP}+\mathrm{TN}}{n} = \frac{\mathrm{P}\cdot\mathrm{TPR}+\mathrm{N}\cdot\mathrm{TNR}}{n} = \frac{\mathrm{P}}{n}\mathrm{TPR}+\frac{\mathrm{N}}{n}\mathrm{TNR}.$
- Accuracy>recall iff TNR>/=/<TPR.
- Accuracy=recall iff _____

Example

D (100)	Predicted +ve (10)	Predicted -ve (90)
Actual +ve (10)	5	5
Actual -ve (90)	5	85

- Accuracy is _____%.
- Precision is _____%.
- NPV is _____%.
- Recall is _____%.
- Specificity is _____%.

Class imbalance problem

- Happens when $\mathrm{P} \ll \mathrm{N}$ (or $\mathrm{N} \ll \mathrm{P}$).
- If $\mathrm{P} \ll \mathrm{N}$, accuracy can be dominated by _____ over _____.

$$\text{Accuracy} = \frac{\mathrm{TP} + \mathrm{TN}}{n} = \frac{\mathrm{P}}{n}\mathrm{TPR} + \frac{\mathrm{N}}{n}\mathrm{TNR} = \frac{\widehat{\mathrm{P}}}{n}\mathrm{PPV} + \frac{\widehat{\mathrm{N}}}{n}\mathrm{NPV}$$

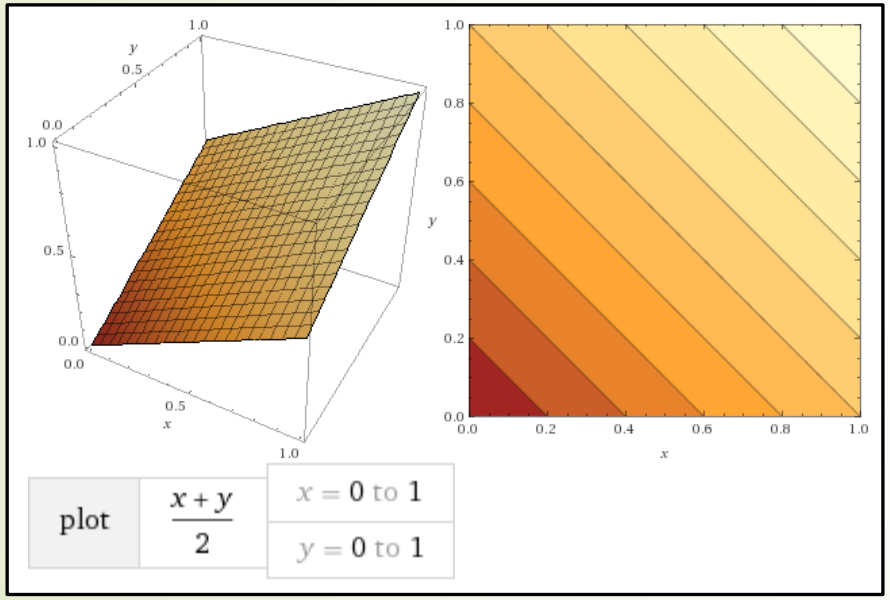
- How to evaluate the prediction of positive class?
- Cost/benefit analysis
  - Different per unit cost/benefit assigned to FP, FN, TP and TN.
  - Minimize total cost or maximize total benefit

$$\text{Cost} = \mathrm{FP} \cdot \mathrm{Cost}_{\mathrm{FP}} + \mathrm{FN} \cdot \mathrm{Cost}_{\mathrm{FN}} + \mathrm{TP} \cdot \mathrm{Cost}_{\mathrm{TP}} + \mathrm{TN} \cdot \mathrm{Cost}_{\mathrm{TN}}$$

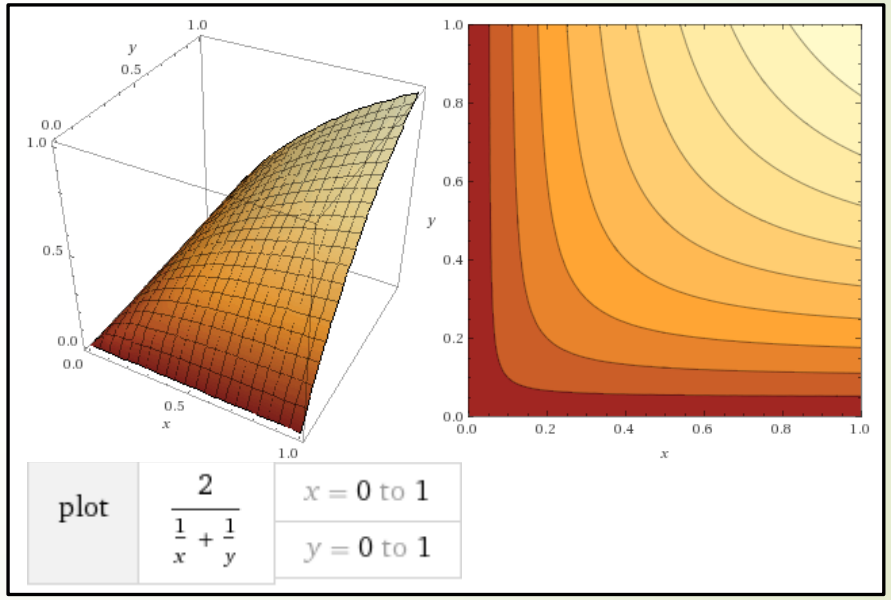
F score

$$F_1 := \left(\frac{\mathrm{PPV}^{-1} + \mathrm{TPR}^{-1}}{2}\right)^{-1} = \frac{2 \cdot \mathrm{PPV} \cdot \mathrm{TPR}}{\mathrm{PPV} + \mathrm{TPR}} \quad (\boldsymbol{F}\textbf{-score/measure})$$

- Why Harmonic means instead of arithmetic mean?



Arithmetic mean=0.7 implies PPV,TPR≥____



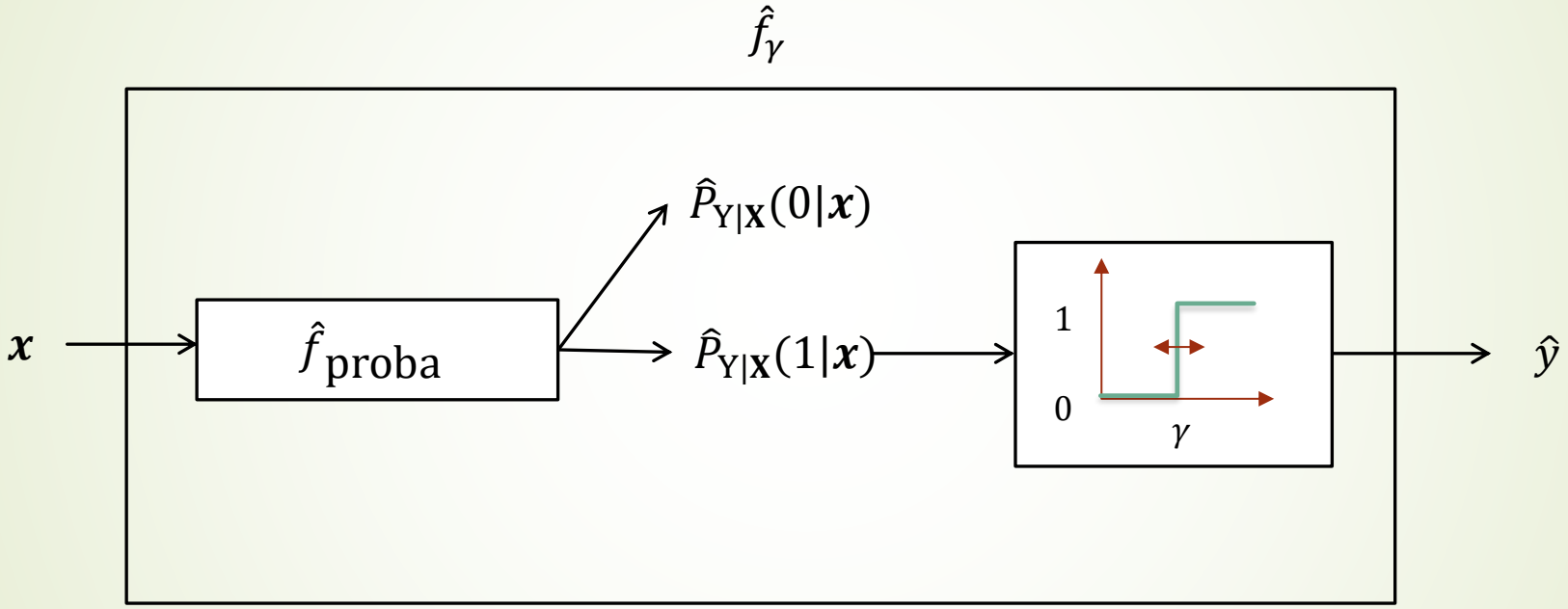
Harmonic mean=0.7 implies PPV,TPR≥_____

F-beta score

$$F_\beta \coloneqq \left(\frac{\mathrm{PPV}^{-1} + \beta^2 \cdot \mathrm{TPR}^{-1}}{\beta^2 + 1}\right)^{-1} = \frac{(\beta^2 + 1) \cdot \mathrm{PPV} \cdot \mathrm{TPR}}{\beta^2 \cdot \mathrm{PPV} + \mathrm{TPR}} \quad \text{for } \beta > 0$$

- As $\beta \to \infty$, $F_\beta \to$____
- As $\beta \to 0$, $F_\beta \to$____

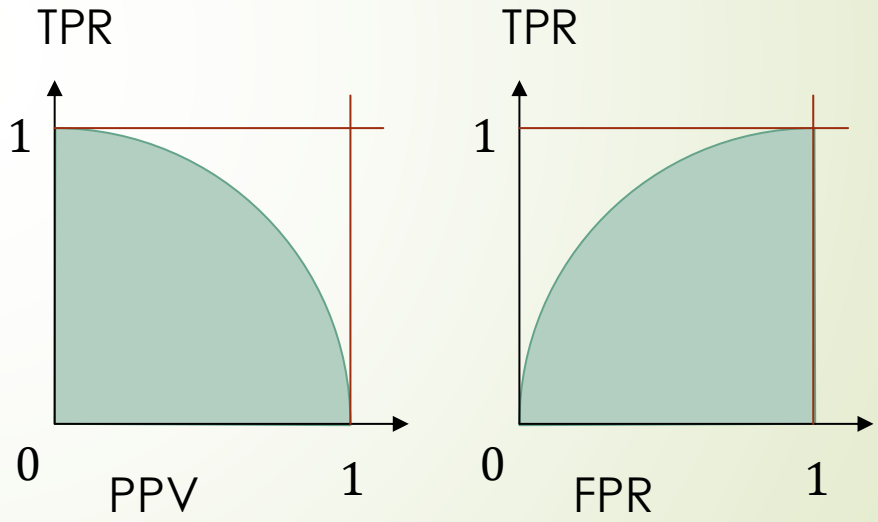
Area under curve (AUC)
Threshold-moving



- Apply a threshold γ to the output of a probabilistic classifier.

Area under curve (AUC)

- Obtain the trade-offs of different performance metrics by varying the threshold.
- **R**eceiver **o**peration **c**haracteristics curve (ROC):
  - Plot of TPR against FPR (False positive rate=1-TNR)
  - AUC: ROC area
- **P**recision **r**ecall curve (PRC):
  - Plot of precision against recall
  - AUC: PRC area
- Which is better, ROC or PRC?



References

- 8.5.1 Metrics for Evaluating Classifier Performance
- 8.5.6 Comparing Classifiers based on Cost-Benefits and ROC Curves.