09.08.2024

Statistical Methods in AI (CS7.403)

Lecture-4: Classification Performance Measures (F1-Score, Multi-Class Metrics)

Ravi Kiran (ravi.kiran@iiit.ac.in)

https://ravika.github.io

@vikataravi

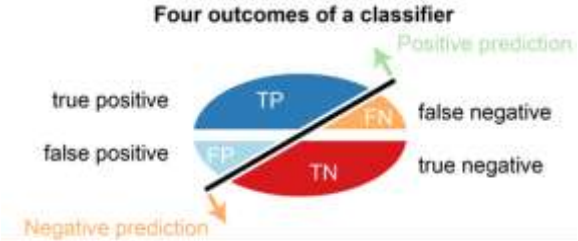
Center for Visual Information Technology (CVIT)

IIIT Hyderabad

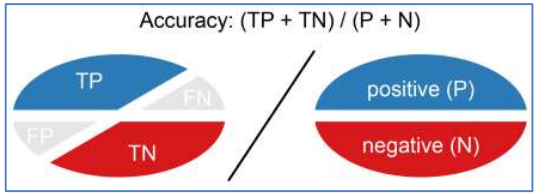
Announcements

- A1
  - Will be out on Saturday (Aug 10)
  - Explanation session in Tutorial (Saturday)

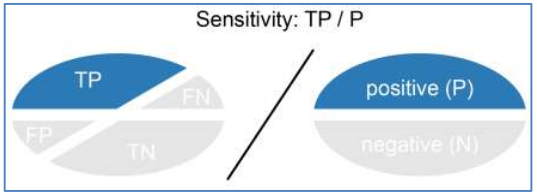
Summary of Measures – Two Class Classification



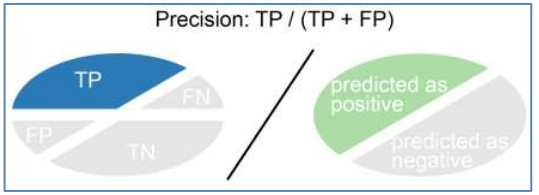
n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	



% of correct predictions



% of + class correctly predicted
[aka Recall / TPR]

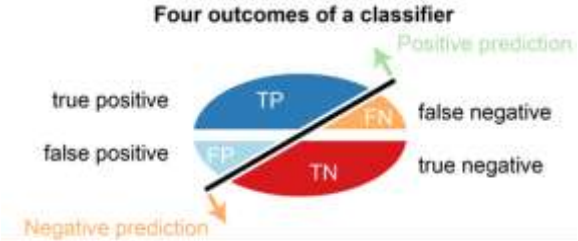


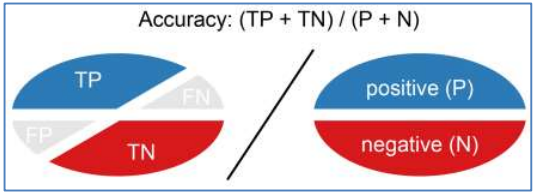
correct prediction of + class

% of – class incorrectly predicted

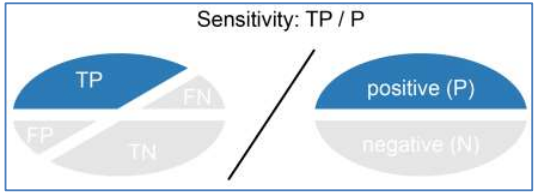
Accuracy vs Precision vs Recall

- Monitor **Precision** if a false positive carries higher cost.
- Monitor **Recall** if a false negative carries higher cost.

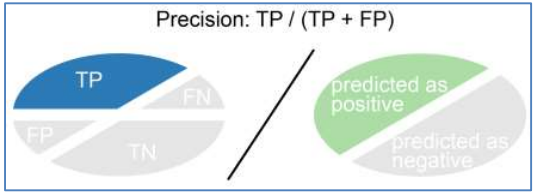




% of correct predictions



% of + class correctly predicted
[aka Recall / TPR]



correct prediction of + class
[aka Precision]

F1-score: A unified measure

- What to do when one classifier has better precision but worse Recall, while other classifier behaves exactly opposite?
  - F-measure (Information Retrieval)
    - $F_1 = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}}$

F-1 score

- What to do when one classifier has better Precision but worse Recall, while other classifier behaves exactly opposite?
  - F-measure (Information Retrieval)
    - $F_1 = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}}$

→ **F1 measure punishes extreme values more !**

→ **Definition of Recall and Precision have same numerator, different denominators. A sensible way to combine them is harmonic mean.**

Classification

Binary

Multi-class

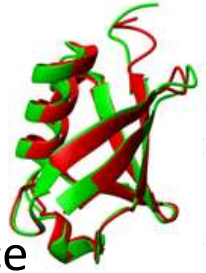
Multi-label

Structure

{0,1}

1-of-K

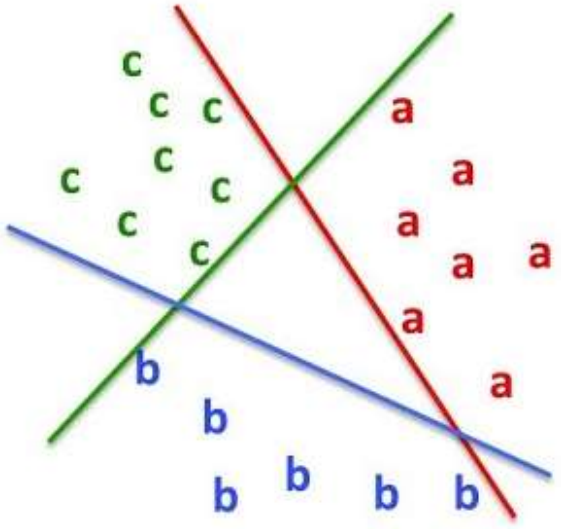
n-of-K



E.g. graph/sequence

How to use 2-class measures for multi-class ?

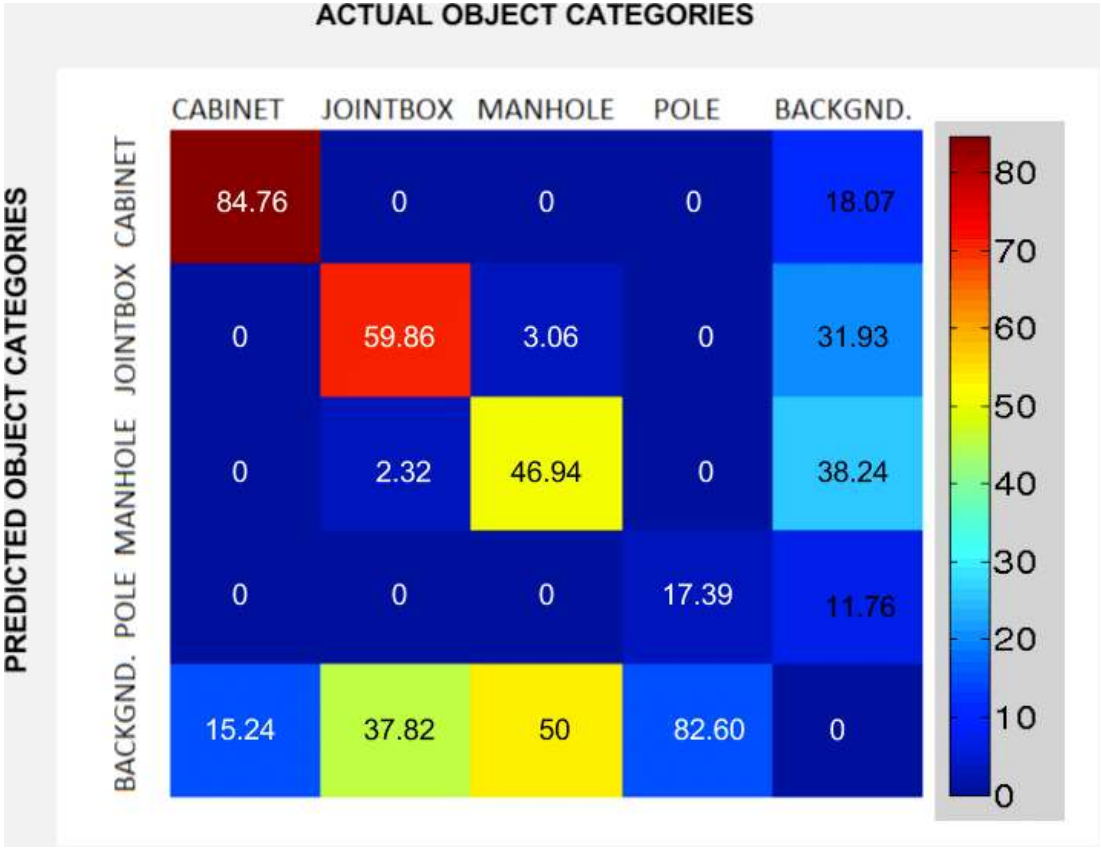
- Convert into 2-class problems !



- Average Precision, Recall etc.
- Board …

Avg. accuracy may not be very meaningful with imbalanced class label distribution

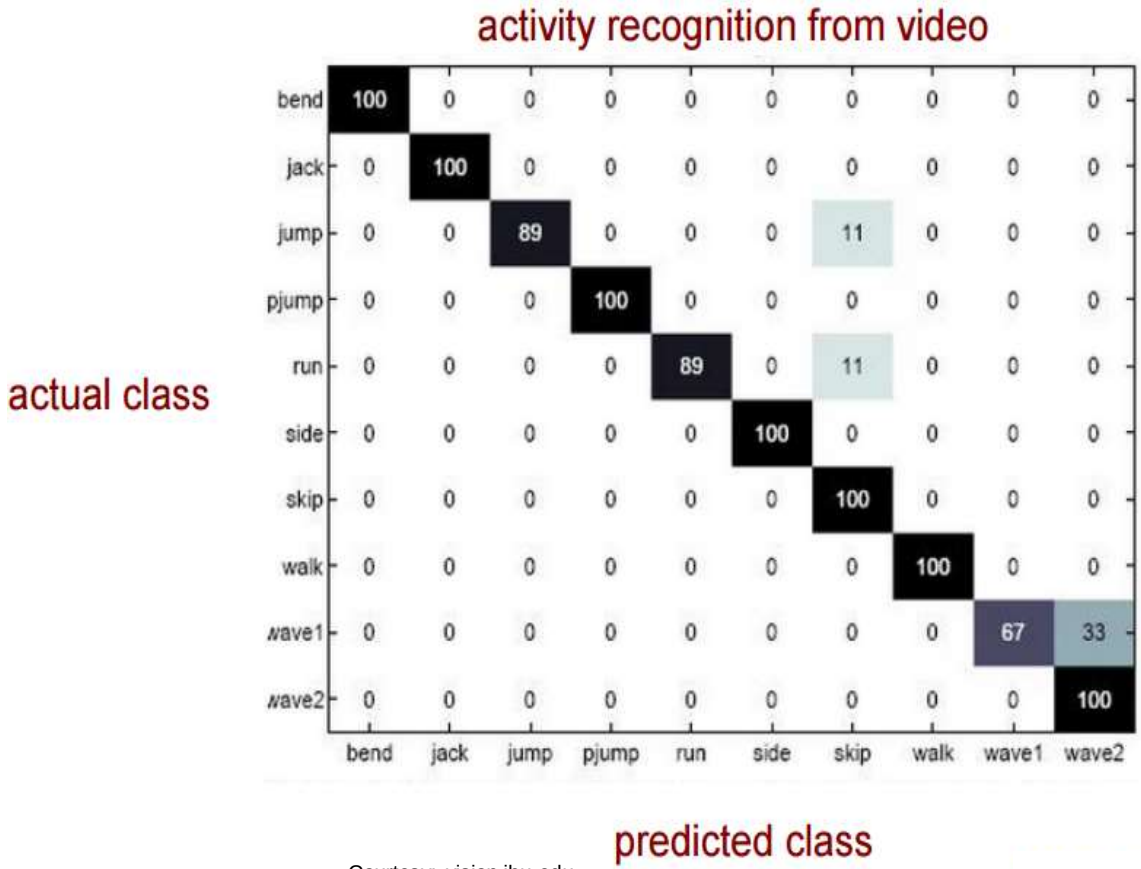
Multi-class Classification - Confusion matrix



- Reveals several performance aspects of the classifier:
  - Most confusing pairs
  - Least confusing pairs

https://www.researchgate.net/figure/Confusion-Matrix-Heat-Map-This-figure-depicts-that-Cabinets-jointboxes-and-manholes-have_fig5_315619273

Multi-class problems - Confusion matrix



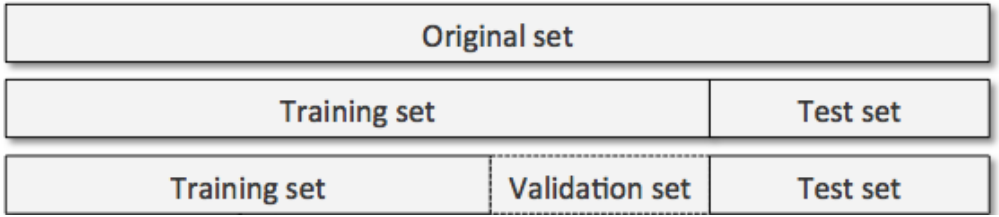
Courtesy: vision.jhu.edu

Multi-class Classification: Measures

- Mean <measure> +- standard deviation
- Median <measure> +- median absolute deviation

Descriptor	Spectral bands	
	RGB	PCA RGB
Gist	74.14 ± 1.93	77.76 ± 2.62
MSIFT	88.92 ± 1.39	90.97 ± 1.81
MBoW	88.60 ± 1.70	88.31 ± 1.38
cSIFT	88.17 ± 1.17	88.76 ± 1.74
rgSIFT	88.24 ± 1.89	87.71 ± 1.33
BoWV [8]	71.86	N/A
SPMK [12]	74.00	N/A
SPCK++ [8]	76.05	N/A
Dense SIFT [2]	81.67 ± 1.23	N/A

Exam analogy: Did you prepare at least a little ?



- Compute <Performance Measure> (e.g. Accuracy) for TRAINING SET
- Verify it is “decent”

Baseline Classifiers

- 0 cost-to-build classifiers
- Binary
  - Equal # of samples / class **in training set** → Random Guessing (50% accuracy)
  - Class imbalance
    - → Guess according to class proportion (Accuracy =)
    - 0-Rule: Majority class (Accuracy =) [slightly stronger baseline]

Classification

Binary

{0,1}

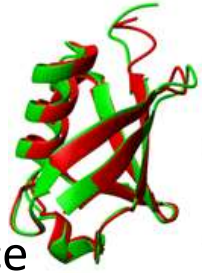
Multi-class

1-of-K

Multi-label

n-of-K

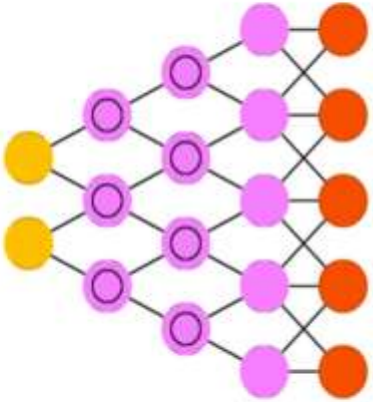
Structure

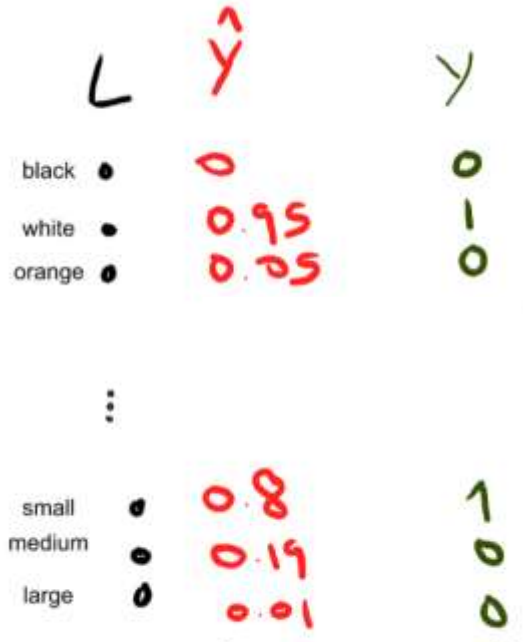


E.g. graph/sequence

Multi-Label Classification

X





Multi Label Classifier

Multi-Label Classifier

https://pyimagesearch.com/2018/05/07/multi-label-classification-with-keras/

Example-based

- n is the number of examples.
- Y_i is the ground truth label assignment of the i^{th} example..
- x_i is the i^{th} example.
- $h(x_i)$ is the predicted labels for the i^{th} example.

Precision = $\frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i \cap h(x_i)|}{|h(x_i)|}$

What % of labels are predicted correctly ?

Recall = $\frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i \cap h(x_i)|}{|Y_i|}$

What % of correct labels were predicted ?

Accuracy = Fraction of samples predicted correctly

Summary

- Many metrics:
  - Accuracy, TP, FP, Precision, Recall, F-1 score
  - Class imbalance and decision-cost imbalance must be taken into account
- Confusion Matrix: Important to analyze and refine solution.

References and Reading

- Code
  - https://scikit-learn.org/stable/modules/model_evaluation.html#classification-metrics