UNC CHARLOTTE

The WILLIAM STATES LEE COLLEGE *of* ENGINEERING

Introduction to ML
Lecture 5: Classifier Evaluation

Hamed Tabkhi
Department of Electrical and Computer Engineering,
University of North Carolina Charlotte (UNCC)
htabkhiv@uncc.edu

UNC CHARLOTTE

Evaluation metrics

- Accuracy, error rate
  - Accuracy is the percent of correct classifications
    - **Accuracy** = Correct Predictions / Total Predictions
  - Error rate is the percent of incorrect classifications
  - Accuracy = 1 – Error rate
- Problems with the accuracy
  - Assumes equal costs for misclassification
  - Assumes relatively uniform class distribution
    - E.g. imbalanced dataset. Consider 95 negative samples and 5 positive samples. Classifying all samples as negative in this case gives 0.95 accuracy score.

Evaluation metrics

	Predicted Y	Predicted N
Actually Y	True Positive	False Negative
Actually N	False Positive	True Negative

Evaluation metrics

True Positive: we correctly detect the class

False Positive: we predict a target class for a negative sample - cause false alarm

	Predicted Y	Predicted N
Actually Y	True Positive	False Negative
Actually N	False Positive	True Negative

UNC CHARLOTTE

Evaluation metrics

True Positive: we correctly detect the class

False Positive: we predict a target class for a negative sample
- Cause false alarm

False Negative: We were not able to predict a correct class for a positive sample
- Can be very bad in many applications

	Predicted Y	Predicted N
Actually Y	True Positive	False Negative
Actually N	False Positive	True Negative

Evaluation metrics

True Positive: we correctly detect the class

False Positive: we predict a target class for a negative sample
- Cause false alarm

False Negative: We were not able to predict a correct class for a positive sample
- Can be very bad in many applications

True Negative?:

	Predicted Y	Predicted N
Actually Y	True Positive	False Negative
Actually N	False Positive	True Negative

Evaluation metrics

recall, **sensitivity**, **hit rate**, or **true positive rate** **(TPR)**

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{P}} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

How much of the real 'Yes' cases are detected? **How well can it detect the condition?**

precision or **positive predictive value** **(PPV)**

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

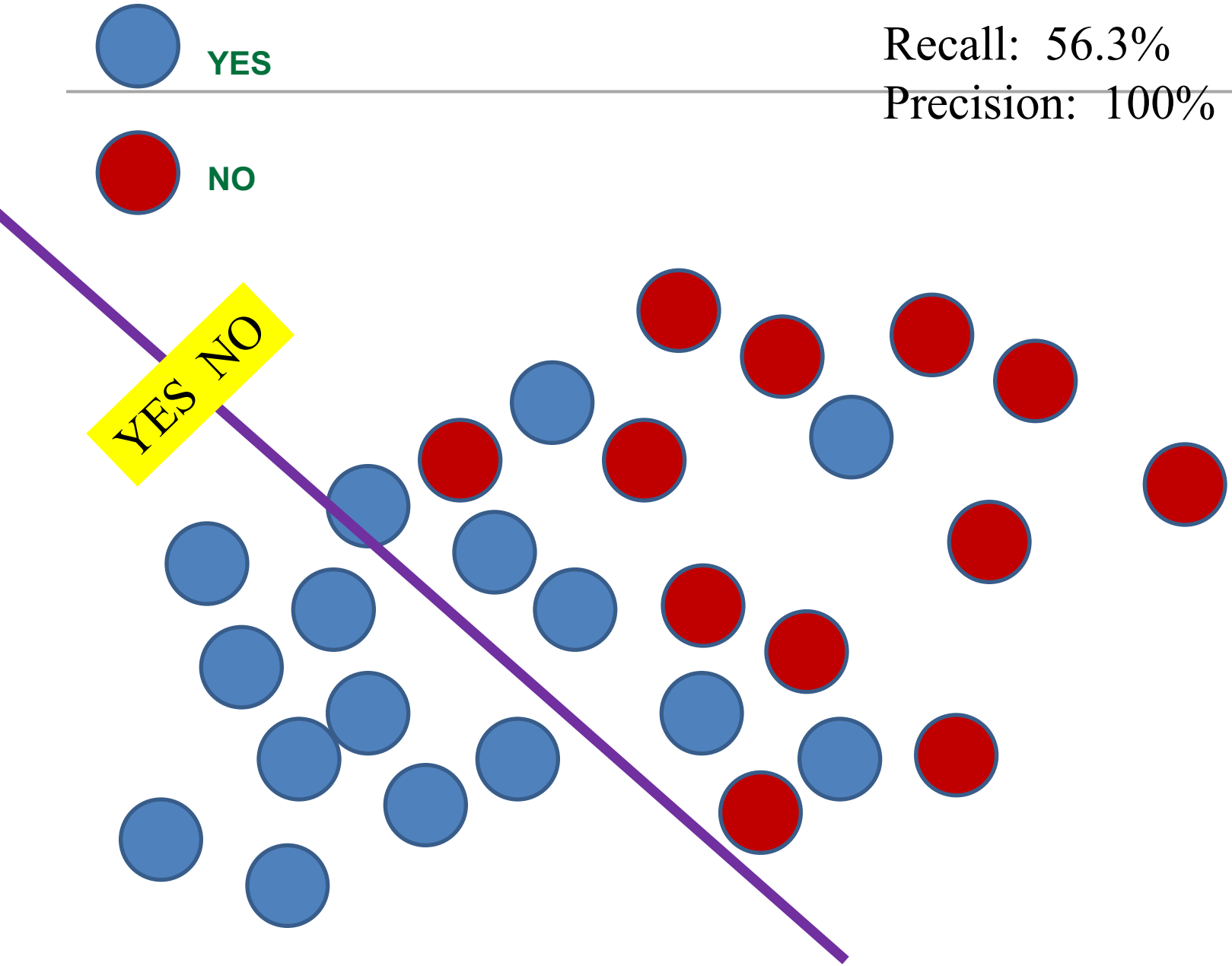
	Predicted Y	Predicted N
Actually Y	True Positive	False Negative
Actually N	False Positive	True Negative

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

Evaluation metrics

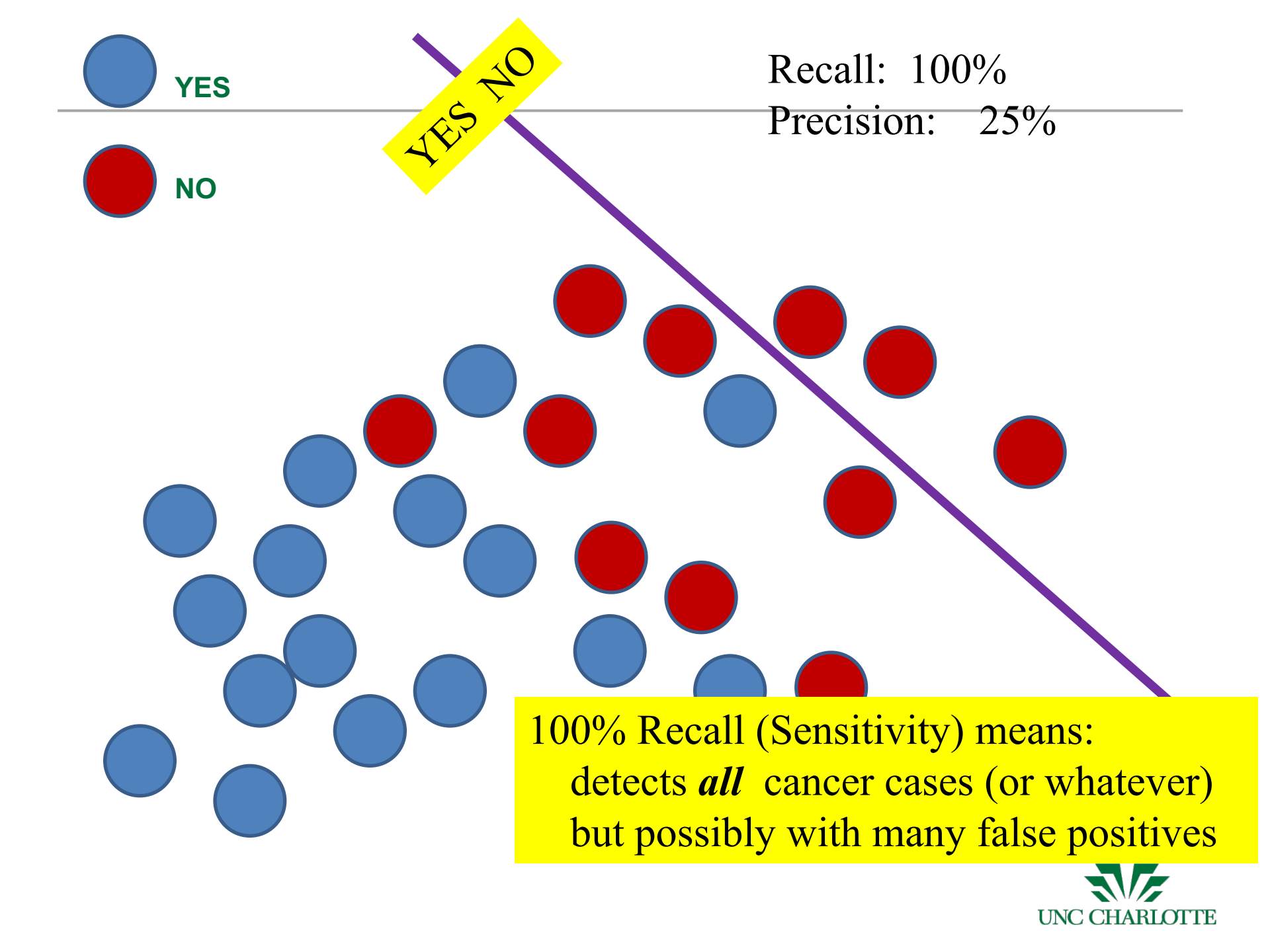
- Previous example: 95 negative samples and 5 positive samples
  - Classifying all samples as negative in this case gives 0.95 accuracy score.

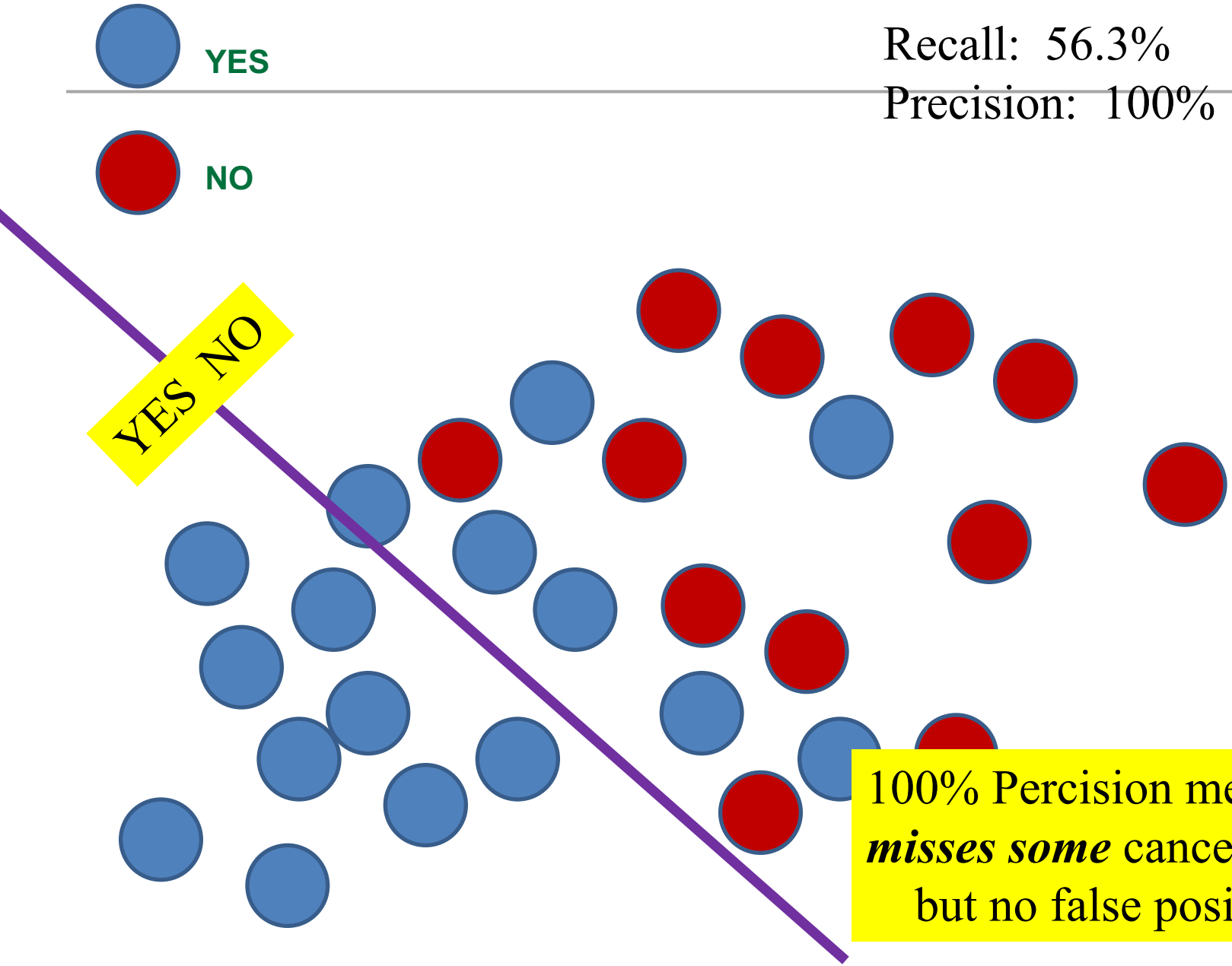
$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn}$$



Recall: 56.3%
Precision: 100%

UNC CHARLOTTE


YES
NO
YES NO
Recall: 100%
Precision: 25%
100% Recall (Sensitivity) means:
detects *all* cancer cases (or whatever)
but possibly with many false positives
UNC CHARLOTTE



Recall: 56.3%
Precision: 100%

100% Percision means:
misses some cancer cases
but no false positives

UNC CHARLOTTE

Evaluation metrics

- Confusion matrix (> 2 classes)

Predicted class

Actual class	1	2	3	4	5	6	7	8	9	Acc
				sum of a corresponding row						
1	**137**	13	3	0	0	1	1	0	0	0.89
2	1	**55**	1	0	0	0	0	6	1	0.86
3	2	4	**84**	0	0	0	1	1	2	0.89
4	3	0	1	**153**	5	2	1	1	1	0.92
5	0	0	3	0	**44**	2	2	1	2	0.82
6	0	0	2	1	4	**35**	0	0	1	0.81
7	0	0	0	0	0	0	**61**	2	2	0.94
8	0	0	0	1	0	0	0	**69**	3	0.95
9	0	0	0	0	0	0	0	2	**26**	0.93
										0.89

What is the total number of test samples of each class?

Evaluation metrics

- Confusion matrix (> 2 classes)

Actual class \ Predicted class	1	2	3	4	5	6	7	8	9	Acc
1	**137**	13	3	0	0	1	1	0	0	0.89
2	1	**55**	1	0	0	0	0	6	1	0.86
3	2	4	**84**	0	0	0	1	1	2	0.89
4	3	0	1	**153**	5	2	1	1	1	0.92
5	0	0	3	0	**44**	2	2	1	2	0.82
6	0	0	2	1	4	**35**	0	0	1	0.81
7	0	0	0	0	0	0	**61**	2	2	0.94
8	0	0	0	1	0	0	0	**69**	3	0.95
9	0	0	0	0	0	0	0	2	**26**	0.93
										0.89

What is the TP for each class?

Each diagonal element corresponds to the TP of a class

Evaluation metrics

- Confusion matrix (> 2 classes)

Actual class \ Predicted class	1	2	3	4	5	6	7	8	9	Acc
1	**137**	13	3	0	0	1	1	0	0	0.89
2	1	**55**	1	0	0	0	0	6	1	0.86
3	2	4	**84**	0	0	0	1	1	2	0.89
4	3	0	1	**153**	5	2	1	1	1	0.92
5	0	0	3	0	**44**	2	2	1	2	0.82
6	0	0	2	1	4	**35**	0	0	1	0.81
7	0	0	0	0	0	0	**61**	2	2	0.94
8	0	0	0	1	0	0	0	**69**	3	0.95
9	0	0	0	0	0	0	0	2	**26**	0.93
										0.89

What is the total number of FN for a class?

The sum of values in the corresponding **row** (excluding the TP)

Evaluation metrics

- Confusion matrix (> 2 classes)

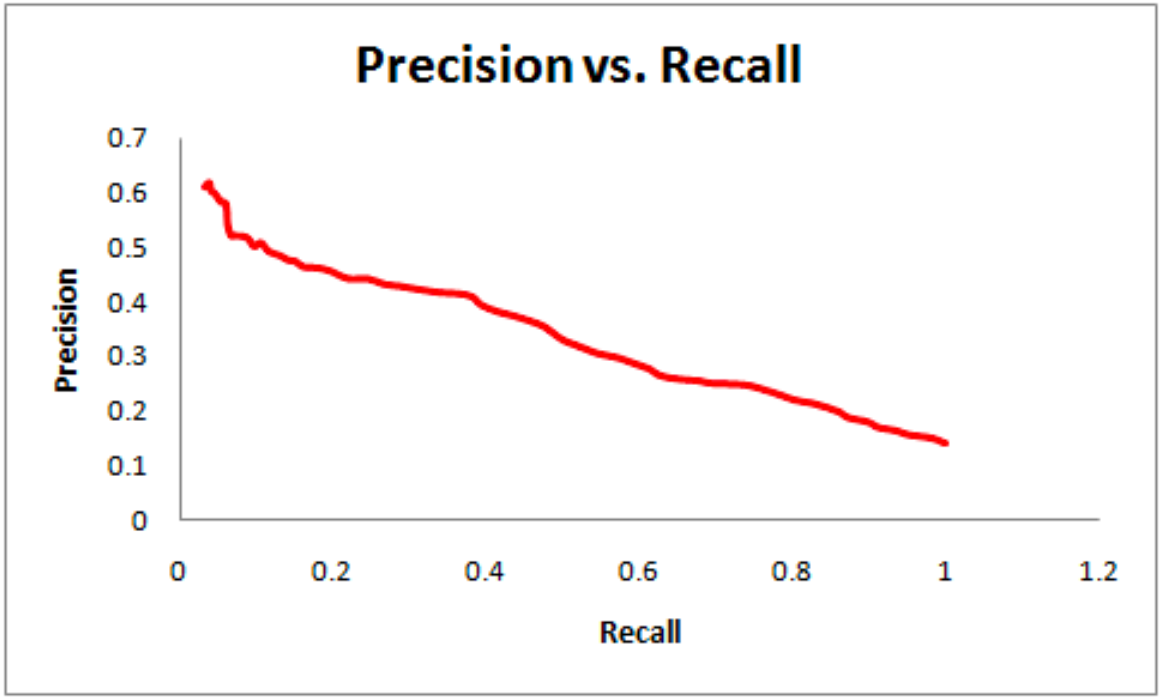
Actual class \ Predicted class	1	2	3	4	5	6	7	8	9	Acc
1	**137**	13	3	0	0	1	1	0	0	0.89
2	1	**55**	1	0	0	0	0	6	1	0.86
3	2	4	**84**	0	0	0	1	1	2	0.89
4	3	0	1	**153**	5	2	1	1	1	0.92
5	0	0	3	0	**44**	2	2	1	2	0.82
6	0	0	2	1	4	**35**	0	0	1	0.81
7	0	0	0	0	0	0	**61**	2	2	0.94
8	0	0	0	1	0	0	0	**69**	3	0.95
9	0	0	0	0	0	0	0	2	**26**	0.93
										0.89

What is the total number of FP for a class?

The sum of values in the corresponding **column** (excluding the TP)

Evaluation metrics

- Precision vs. Recall
  - In practice, one always needs to make a compromise between these two metrics: by increasing Recall, we decrease (though unwillingly) Precision, and vice versa



https://groups.csail.mit.edu/cb/struct2net/webserver/images/prec-v-recall-v2.png

F1 Score

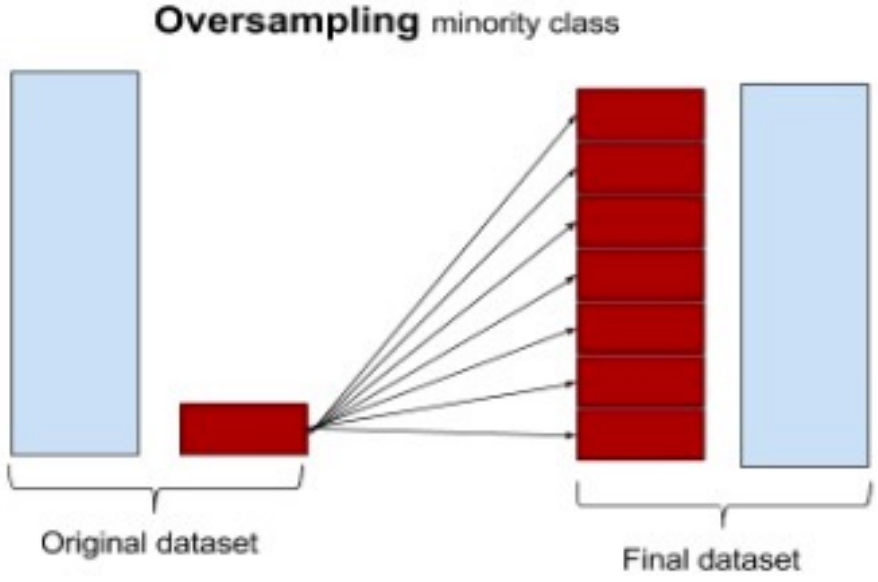
$$= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

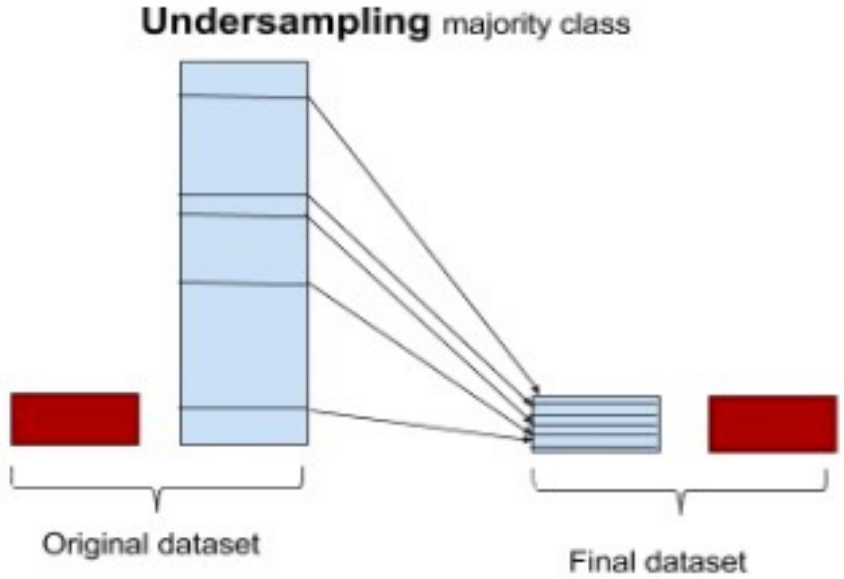
- A higher F1 score means higher accuracy
- It can never be larger than 1

Imbalanced data?

- Solutions
  - Oversampling: re-sampling of data from minority class
  - Under-sampling: randomly eliminate samples from majority class
  - Synthesizing new data points for minority class
    - Take averages of samples in minority class
    - Add small noise to samples in minority class
  - We will talk about this more in deep learning

Imbalanced data?





https://www.svds.com/learning-imbalanced-classes/

UNC CHARLOTTE