AI-503 Advanced Machine Learning

Last time

Feature Scaling

Regularization in SVMs

- Via the hyperparameter C
- Any other factor affecting regularization in SVMs?
  - Kernel Parameters

Evaluation Metrics for classification

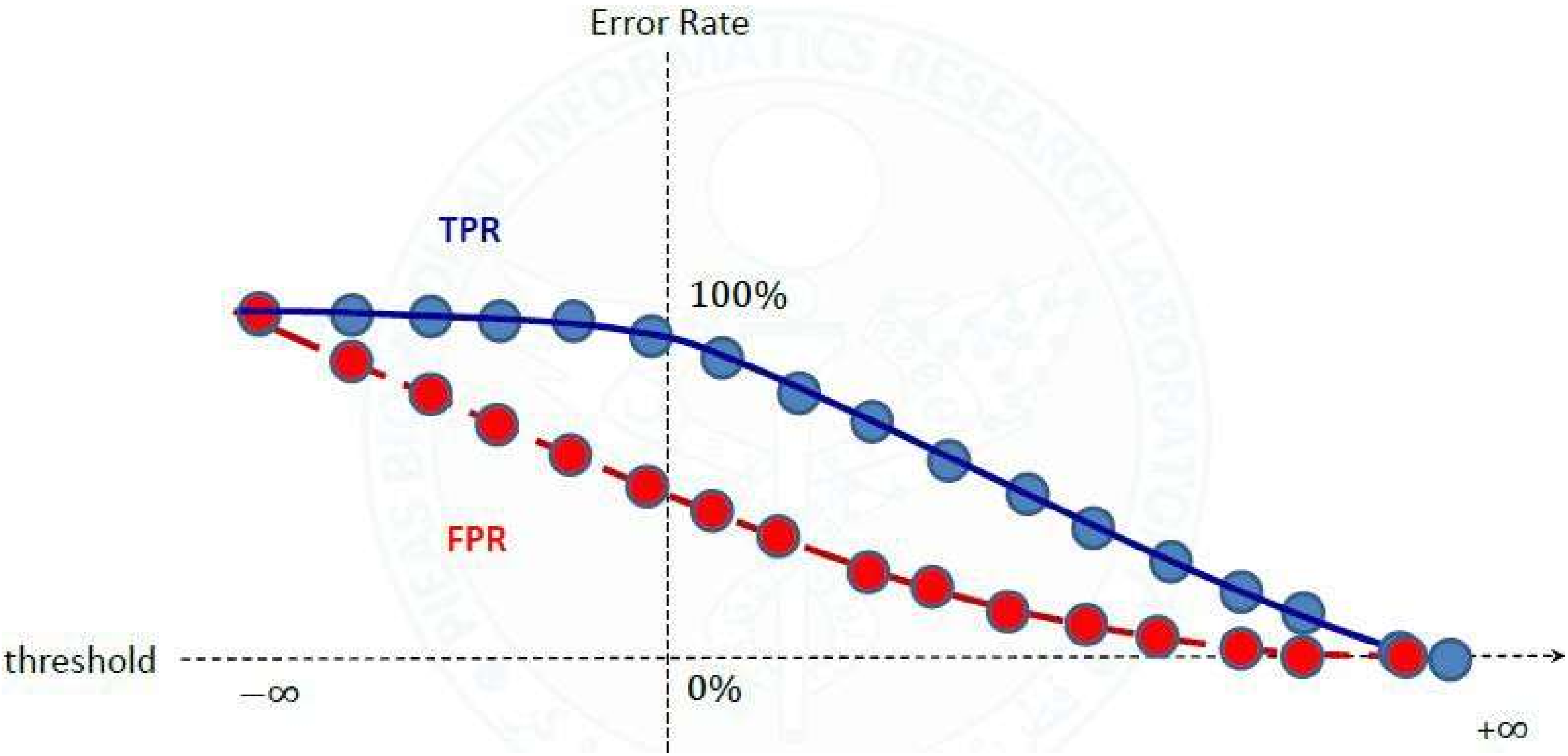
- Why we need them
  - Hyperparameter tuning
  - Model Evaluation
- Accuracy, ROC, AUC-ROC

Name	Formula
error	$(fp + fn)/N$
accuracy	$(tp + tn)/N = 1 - \text{error}$
tp-rate	tp/p
fp-rate	fp/n
precision	tp/p'
recall	$tp/p = \text{tp-rate}$
sensitivity	$tp/p = \text{tp-rate}$
specificity	$tn/n = 1 - \text{fp-rate}$

Behavior of metrics

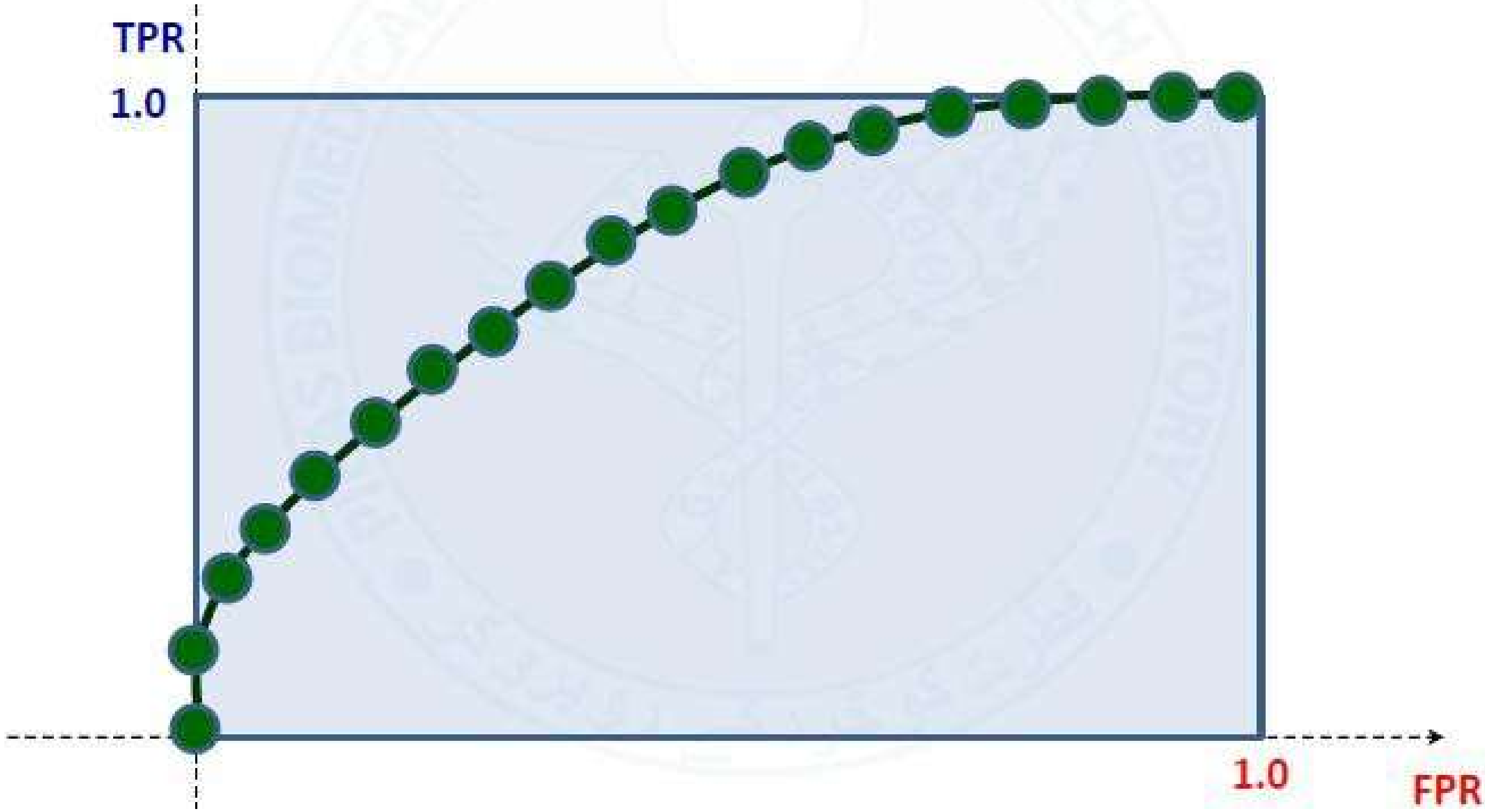
- What will be the behavior of TPR with increase in threshold of the classifier?
- How will FPR behave?
- How will Precision behave?
- Can TPR decrease with increase in threshold?

FPR vs. TPR Curve



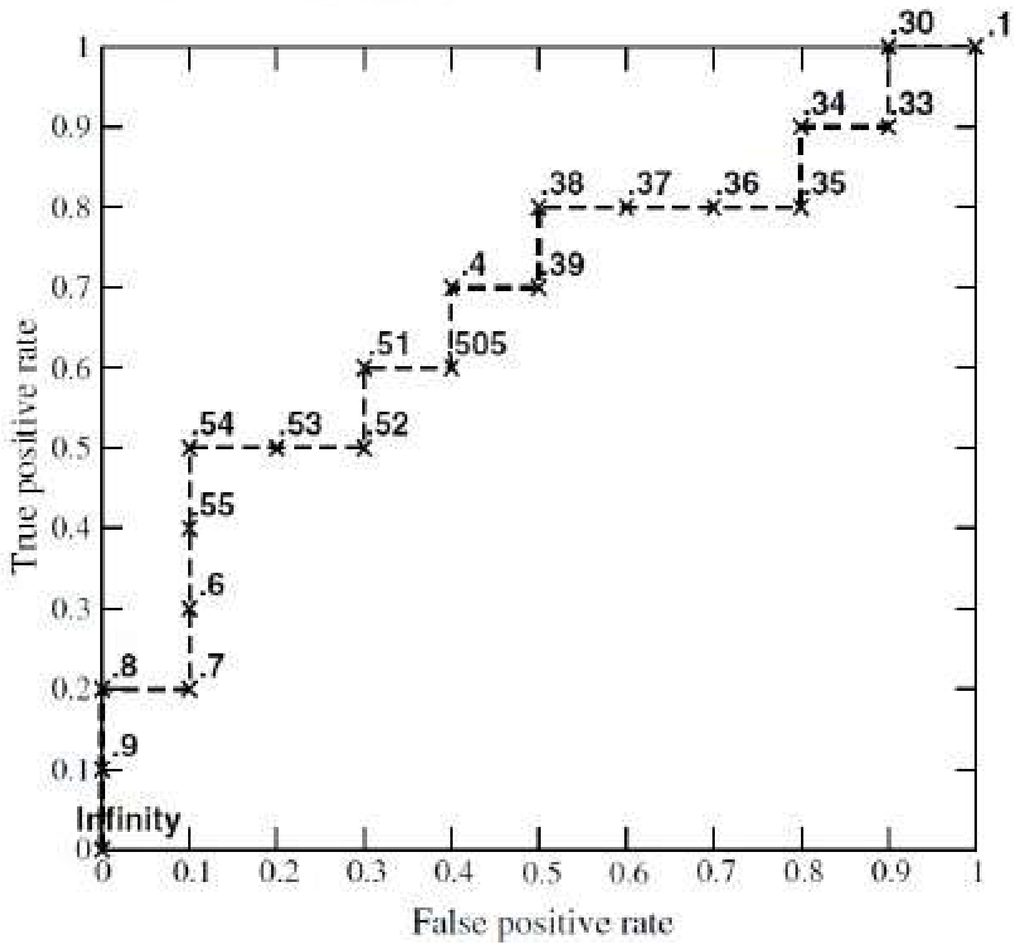
Receiver Operating Characteristics Curve

- A plot of TPR vs FPR



Making the ROC Curve

Inst#	Class	Score	Inst#	Class	Score
1	p	.9	11	p	.4
2	p	.8	12	n	.39
3	n	.7	13	p	.38
4	p	.6	14	n	.37
5	p	.55	15	n	.36
6	p	.54	16	n	.35
7	n	.53	17	p	.34
8	n	.52	18	n	.33
9	p	.51	19	p	.30
10	n	.505	20	n	.1



Example

Instance	Actual Label (y)	Predicted Score (s)
A	1	0.9
B	0	0.8
C	1	0.7
D	0	0.6
E	1	0.55
F	0	0.4

Example

Sort using Score

Instance	Label	Score
A	1	0.9
B	0	0.8
C	1	0.7
D	0	0.6
E	1	0.55
F	0	0.4

Example

Threshold ≥	Predicted Positives	TP	FP	TPR (Recall)	FPR
> 0.9	none	0	0	0 / 3 = 0.00	0.00
≥ 0.9	A	1	0	1 / 3 = 0.33	0.00
≥ 0.8	A, B	1	1	1 / 3 = 0.33	1 / 3 = 0.33
≥ 0.7	A, B, C	2	1	2 / 3 = 0.67	1 / 3 = 0.33
≥ 0.6	A, B, C, D	2	2	2 / 3 = 0.67	2 / 3 = 0.67
≥ 0.55	A, B, C, D, E	3	2	3 / 3 = 1.00	2 / 3 = 0.67
≥ 0.4	All	3	3	3 / 3 = 1.00	3 / 3 = 1.00

Plot the ROC Points

(0.00, 0.00)
(0.00, 0.33)
(0.33, 0.33)
(0.33, 0.67)
(0.67, 0.67)
(0.67, 1.00)
(1.00, 1.00)

The trapezoidal rule

The trapezoidal rule formula between two points is

$$\text{Area} = \frac{(y_i + y_{i+1})}{2} \times (x_{i+1} - x_i)$$

Plot the ROC Points

(0.00, 0.00)
(0.00, 0.33) = 0
(0.33, 0.33) = 0.1089
(0.33, 0.67) = 0
(0.67, 0.67) = 0.2278
(0.67, 1.00) = 0
(1.00, 1.00) = 0.33
Total AUC = 0.6667

ROC curves

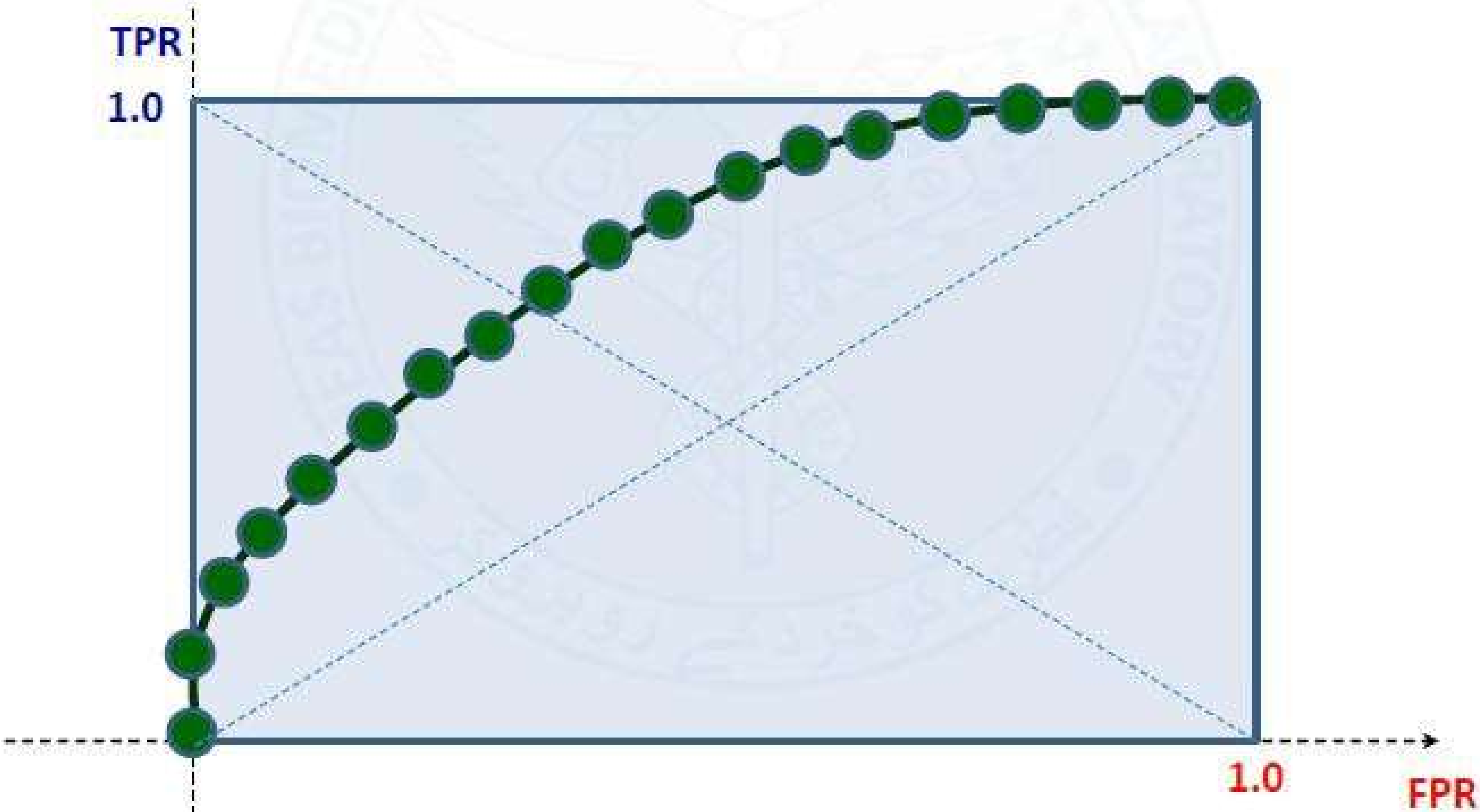
- What will be ROC curve for a perfect classifier?
- What will the ROC Curve of a random classifier look like?
- What will the ROC curve of a classifier that always predicts the positive class look like?
- What are the underlying assumptions of the ROC curve?
- What part of the ROC curve is the most important?

ROC Assumptions

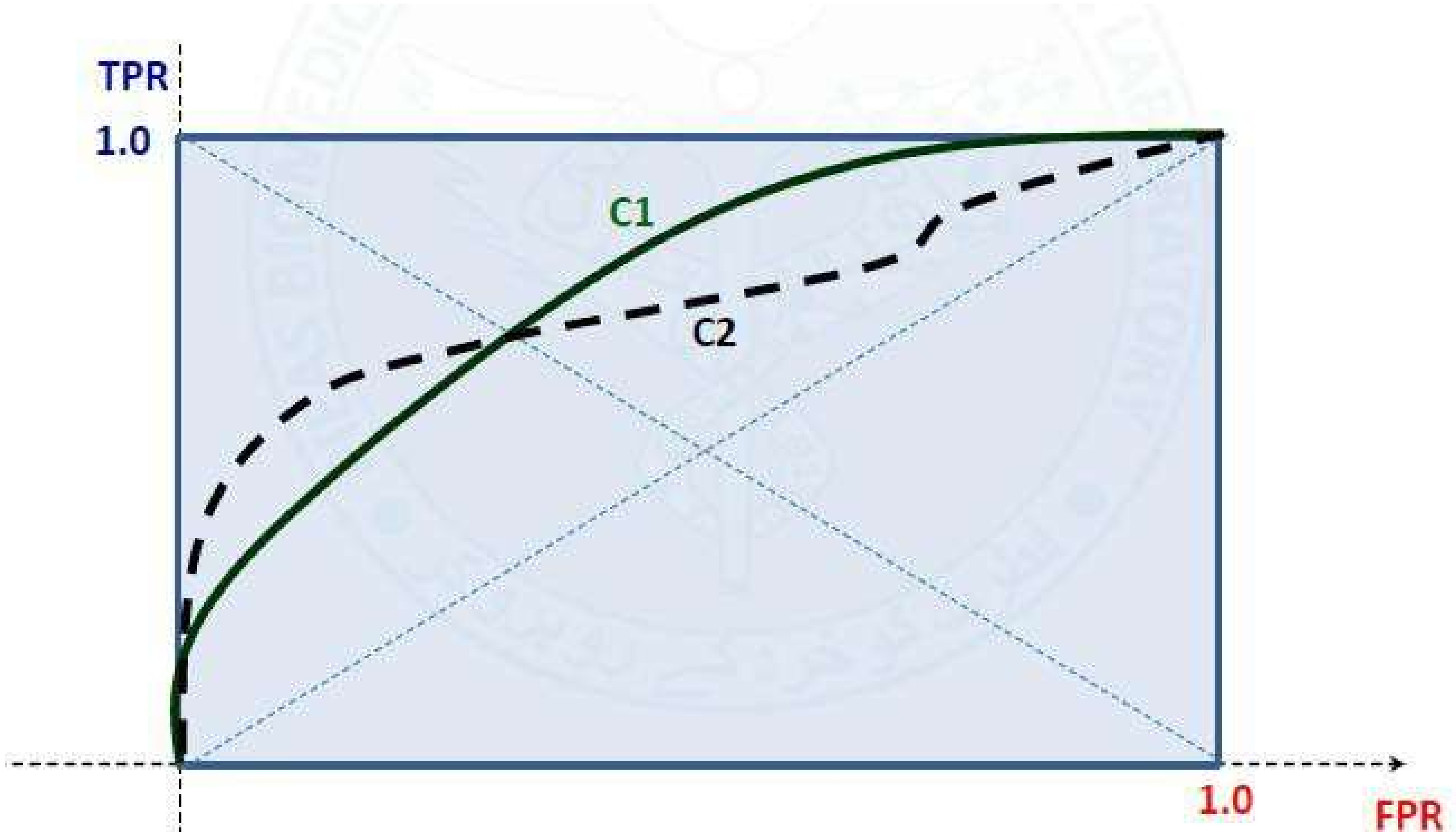
- Binary classification — Only two classes (positive and negative)
- Classifier outputs real-valued scores — Not just hard labels
- Thresholds can be varied — To evaluate different TPR and FPR
- True labels are known and reliable
- Class balance doesn't affect the curve — Unlike precision-recall curves

AUC-ROC

- The area under the ROC curve is a quality metric

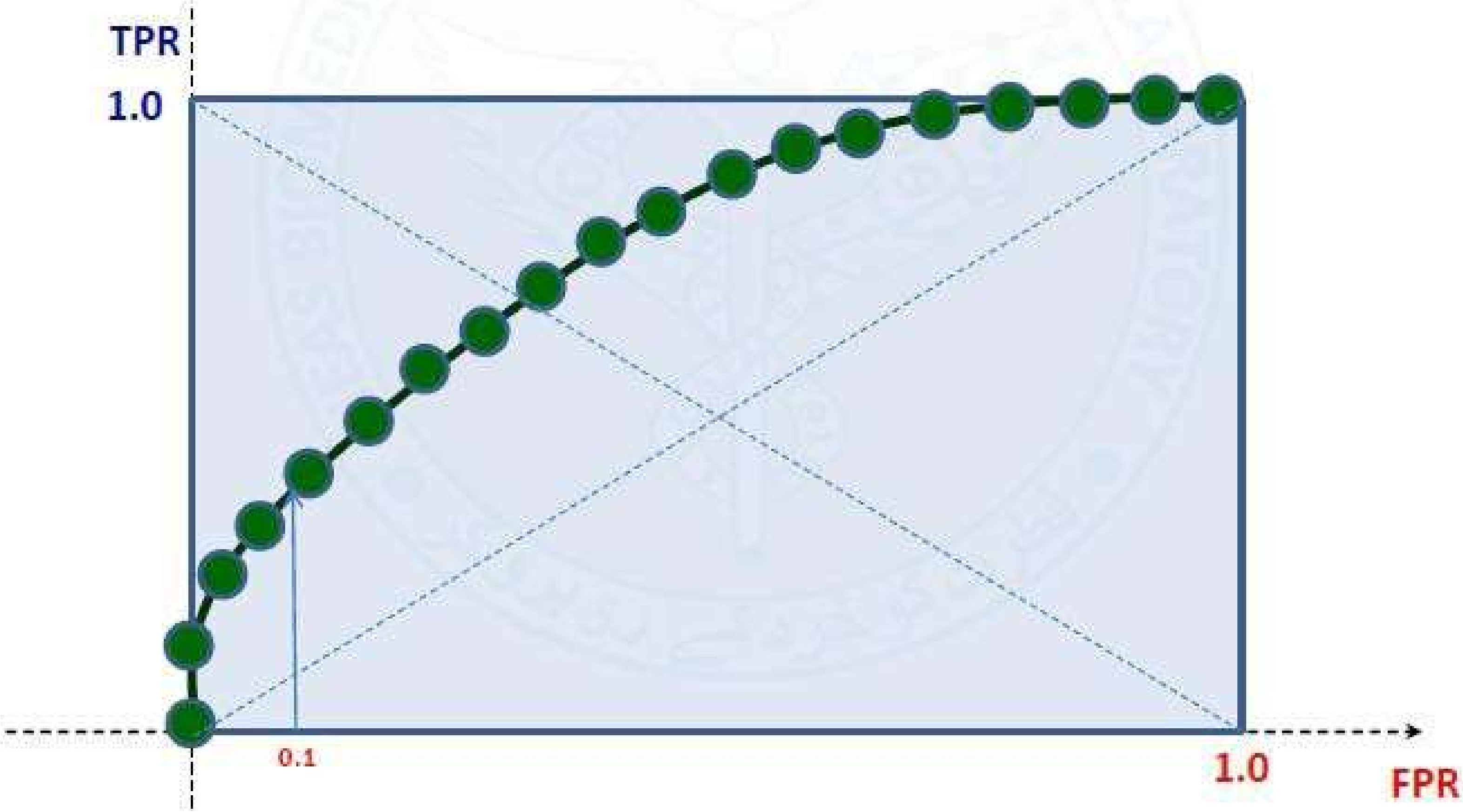


Which one is better?



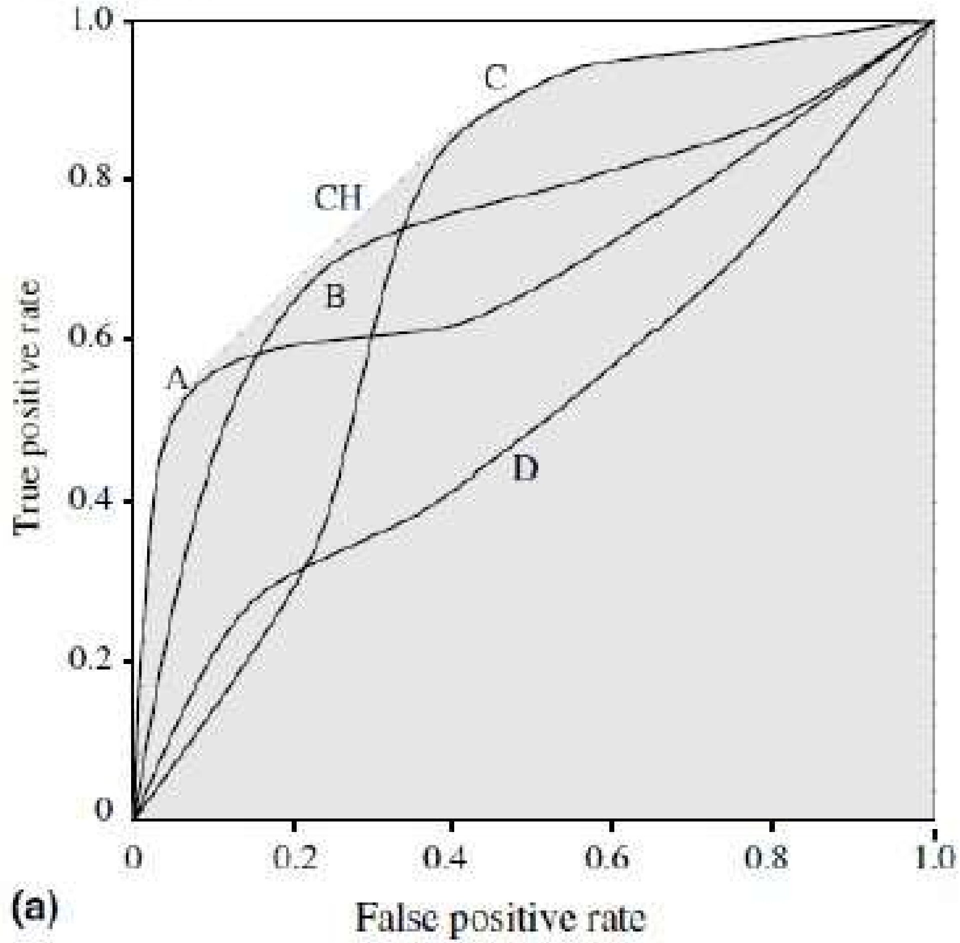
AUC ROC-N

- Area under the ROC curve up to the first N False Positives
  - N = 50
  - N = 10%



ROC Convex Hull

- Scores of two classifiers can be combined through a weighted combination to result in an optimal classifier
- This can be done using the ROC convex hull



(a)

ROC Convex Hull

Construction of the ROC Convex Hull

1. **Plot all ROC points** for a classifier (or multiple classifiers).

2. **Sort points** by increasing FPR (and decreasing TPR for ties).

3. **Connect non-dominated points** with straight lines to form the upper boundary (convex hull).

Multi-class ROC Curves

- Can also make multiple class ROC curves
  - One vs. Rest
- AUC-ROC can also be computed
  - Pairwise

Third type is **averaged ROC** Curves
After building many OvR or OvO curves,
You **average** the metrics (like AUC) across classes.
There are two main types:
Macro-average: Treat all classes equally (simple average).
Micro-average: Aggregate contributions of all classes based on total counts (better when class imbalance exists).

Properties

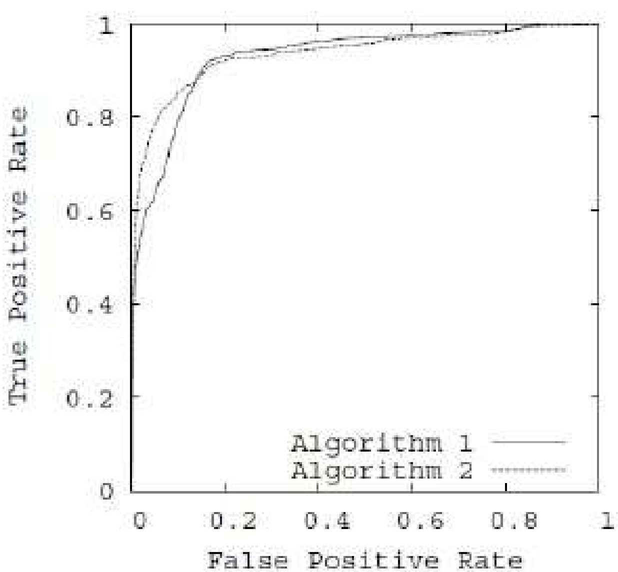
- Class Imbalance?
- When to Use?
- What to focus on?
  - FPR?
  - TPR?

Class imbalance…

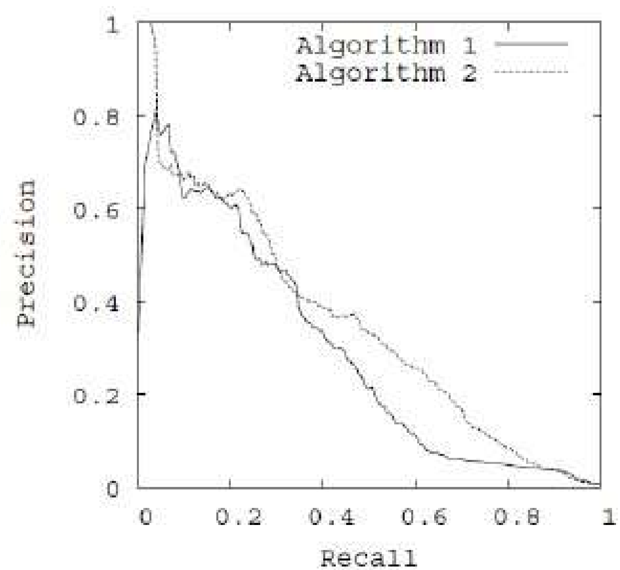
Consider a detector with TP=9, FN=1, TN=900, FP=90, where there are 10 positive and 990 negative sample. TPR=0.9, FPR=0.1 which indicates a good ROC score, however Precision=0.1 which indicates a bad PR score.

Precision Recall Curve

- Plot of Precision vs. Recall
- AUC-PR is a performance metric
- Useful in cases of class-imbalance or in which precision is a requirement



(a) Comparison in ROC space



(b) Comparison in PR space

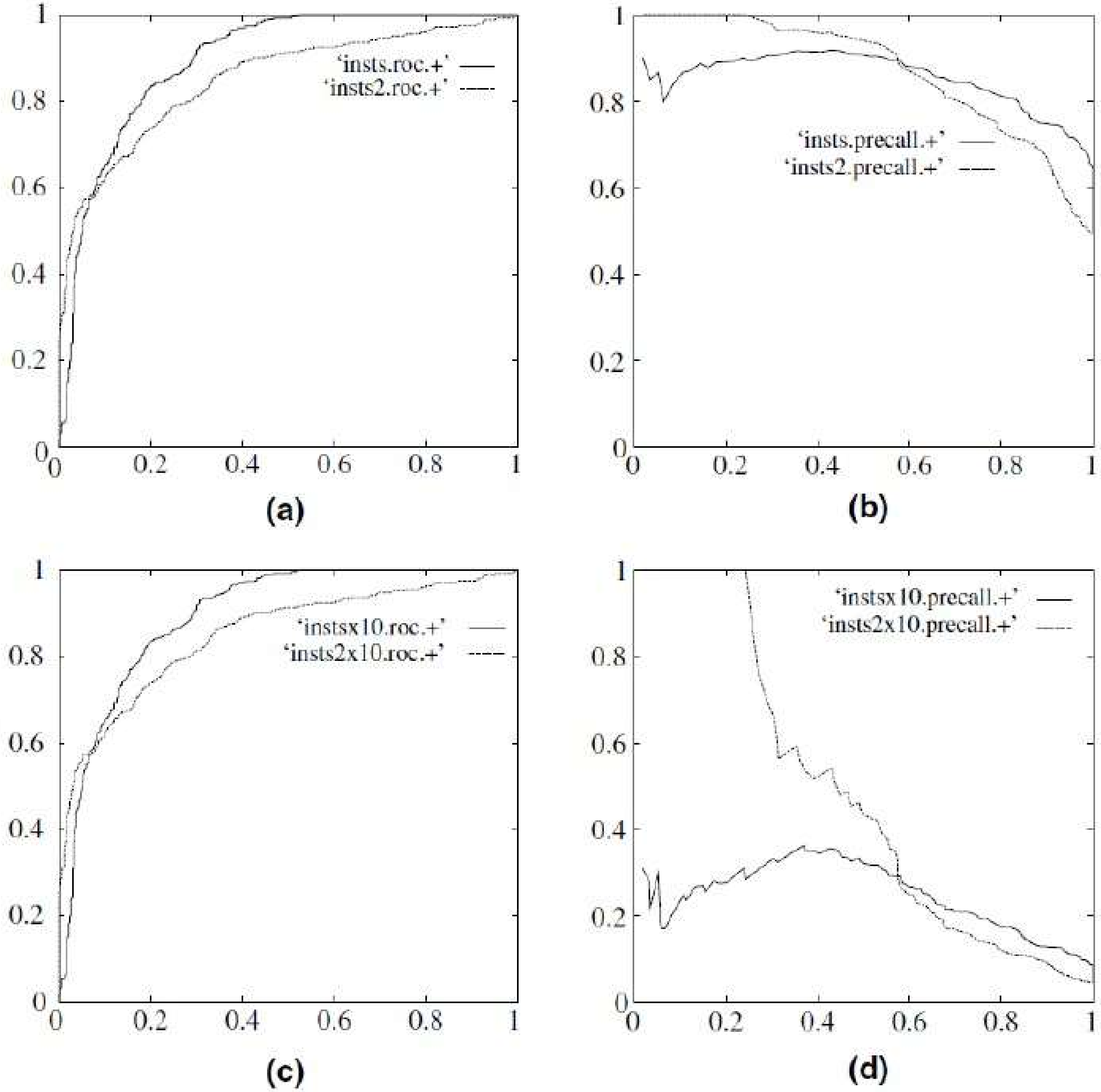
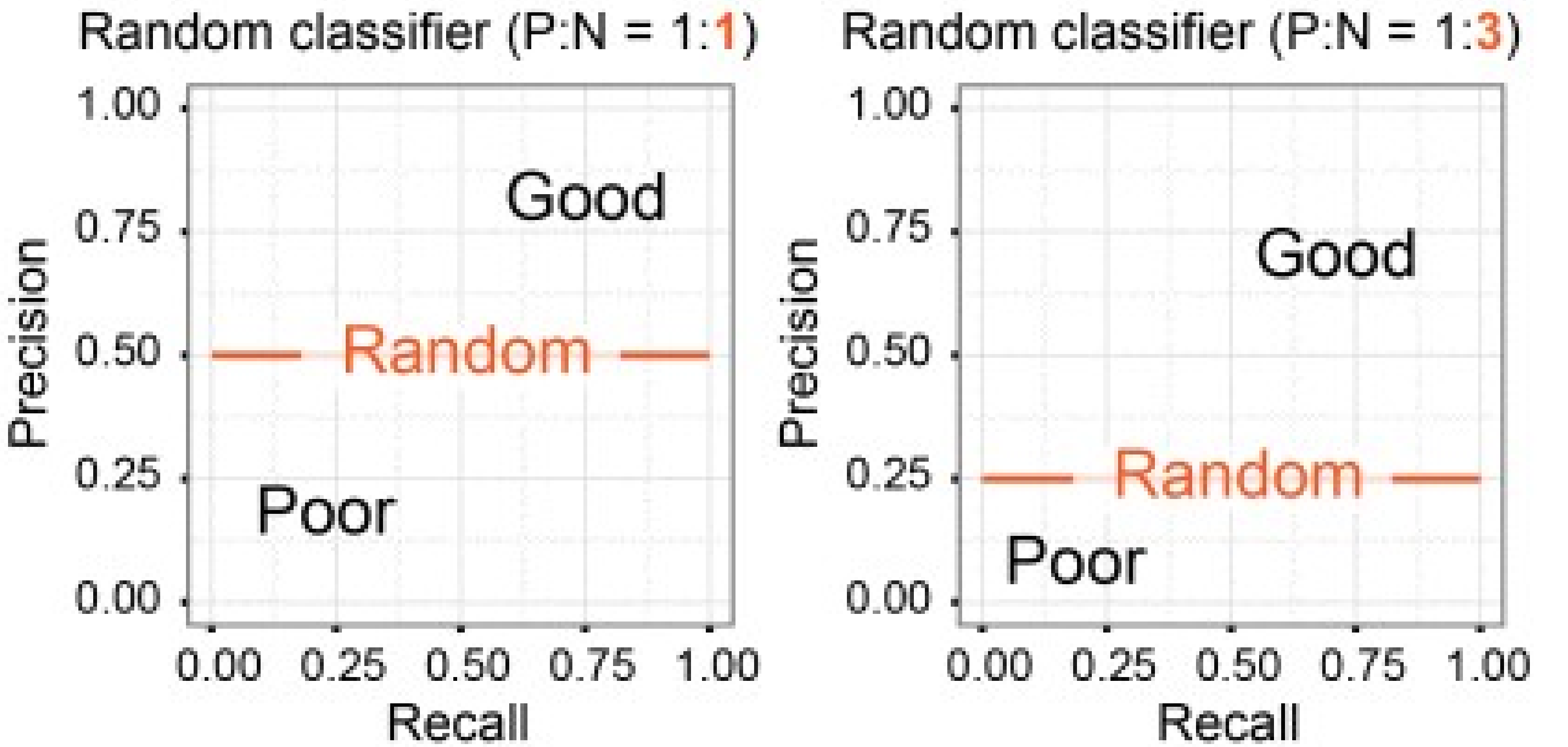


Fig. 5. ROC and precision-recall curves under class skew. (a) ROC curves, 1:1; (b) precision-recall curves, 1:1; (c) ROC curves, 1:10 and (d) precision-recall curves, 1:10.

Relationship between ROC & PR Curves

- One-to-One correspondence between the two curves
- If a curve dominates in ROC space then it dominates in PR space.
- If a curve dominates in PR space then it dominates in ROC space.
- What will be the PR curve for a random classifier?



Random classifier (P:N = 1:1)
Good
Random
Poor
Precision
Recall
0.00
0.25
0.50
0.75
1.00
Random classifier (P:N = 1:3)
Good
Random
Poor
Precision
Recall

ROC and PR Curves in Scikit-Learn

- http://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_curve.html
- http://scikit-learn.org/stable/auto_examples/model_selection/plot_roc.html#example-model-selection-plot-roc-py

- **`from sklearn.metrics import *`**
- `P,R = precision_recall_curve(Y,Z)`
- `AUCPR = average_precision_score(Y,Z)`
- `roc_curve, auc`

Demo. . .

More Scikit Metrics

- http://scikit-learn.org/stable/modules/model_evaluation.html
- F-measure
- Mathews Correlation Coefficient
- Confusion Matrix
- Multiclass metrics

- Read them when you need them!

Reading

- Recommended
  - Davis, Jesse, and Mark Goadrich. 2006. "The Relationship Between Precision-Recall and ROC Curves." In *Proceedings of the 23rd International Conference on Machine Learning*, 233–40. ICML '06. New York, NY, USA: ACM. doi:10.1145/1143844.1143874.
  - Fawcett, Tom. 2006. "An Introduction to ROC Analysis." *Pattern Recogn. Lett.* 27 (8): 861–74. doi:10.1016/j.patrec.2005.10.010.

- Required
  - Alpaydin 2010, Section 19.7