COMS 4030A/7047A Adaptive Computation and Machine Learning

Hima Vadapalli

Semester I, 2022

So far:
Supervised and Unsupervised Learning

Today:
Error Analysis(Evaluation Metrics)

Slides based heavily on course material by Yining Chen, Anand Avati and Andrew Ng

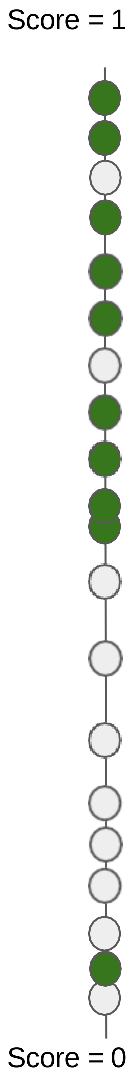
Why are metrics important?

- Training objective (cost function) is only a proxy for real world objectives
- Metrics help capture a business goal into a quantitative target (not all errors are equal).
- Helps organize ML team effort towards that target.
    - Generally in the form of improving that metric on the dev set.
- Useful to quantify the “gap” between:
    - Desired performance and baseline (estimate effort initially).
    - Desired performance and current performance.
    - Measure progress over time.

Binary Classification

- x is input
- y is binary output (0/1)
- Model is $\hat{y} = h(x)$
- Two types of models
  - Models that output a categorical class directly (K-nearest neighbor)
  - Models that output a real valued score (Logistic Regression)
    - Score could be margin or probability
    - Need to pick a threshold

Score based models

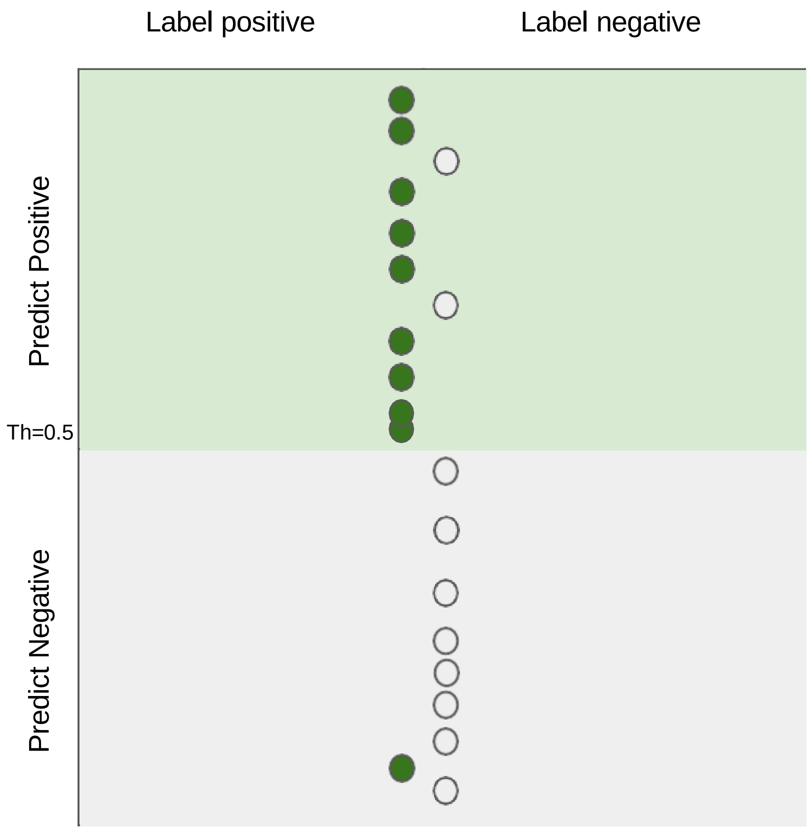


	Positive example
	Negative example

$$\text{Prevalence} = \frac{\#\text{ positive examples}}{\#\text{ positive examples} + \#\text{ negatives examples}}$$

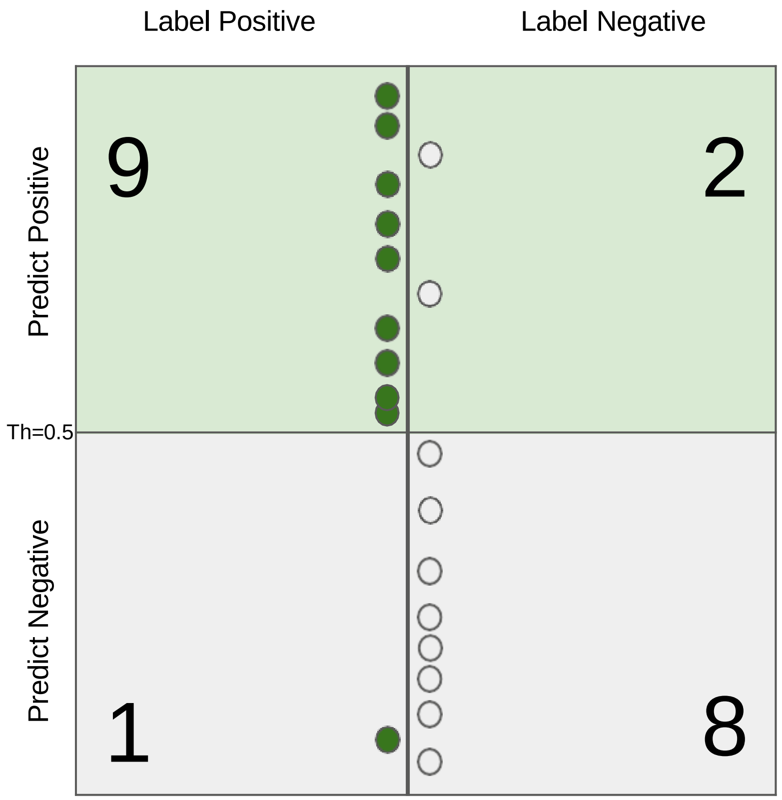
Prevalence lets us understand class imbalance

Threshold -> Classifier -> Point Metrics



Th
0.5

Point metrics: Confusion Matrix

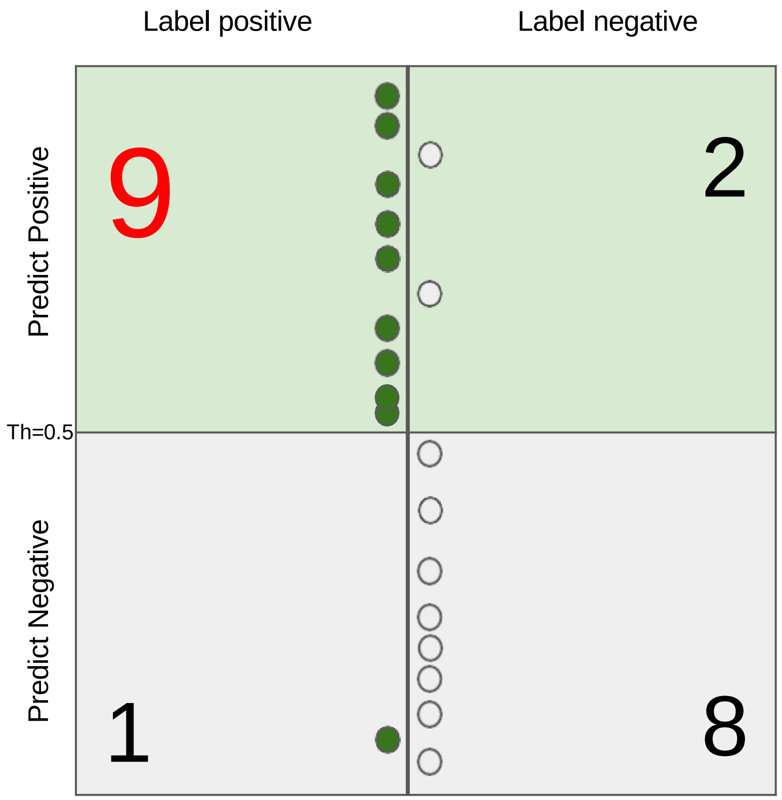


Th
0.5

Properties:

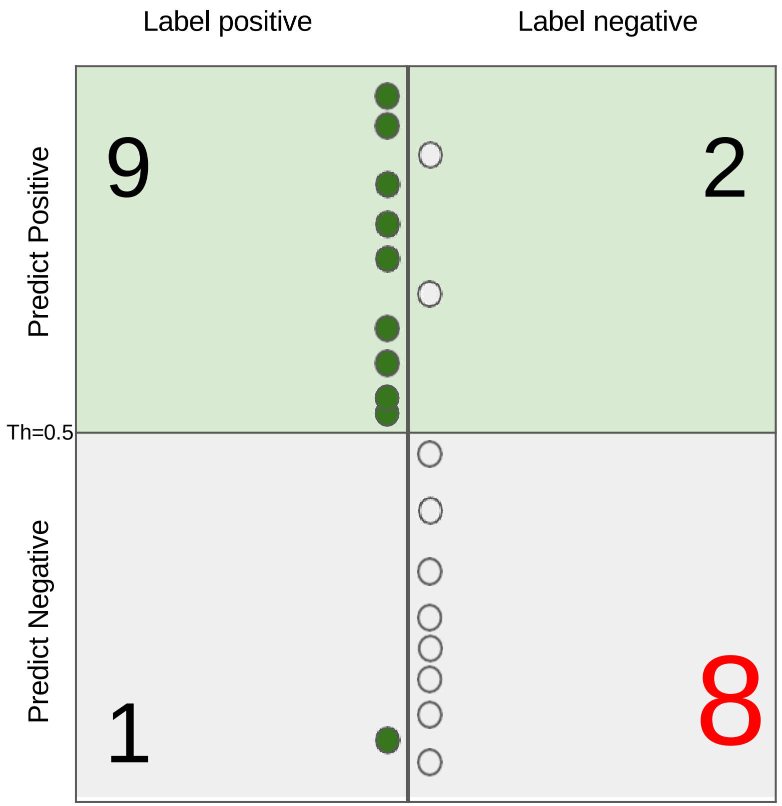
- Total sum is fixed (population).
- Column sums are fixed (class-wise population).
- Quality of model & threshold decide how columns are split into rows.
- We want diagonals to be “heavy”, off diagonals to be “light”.

Point metrics: True Positives



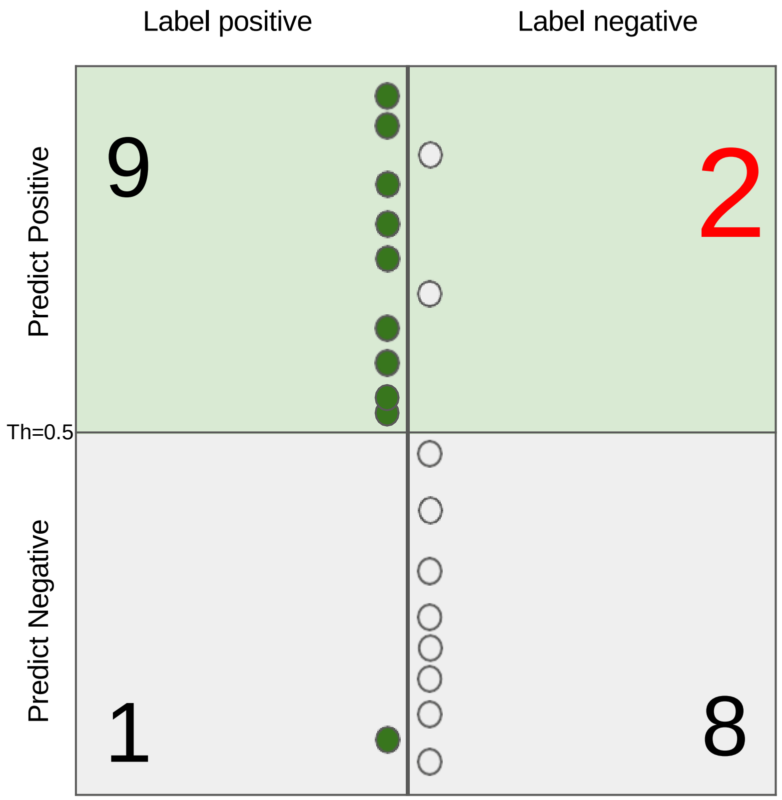
Th	TP
0.5	9

Point metrics: True Negatives



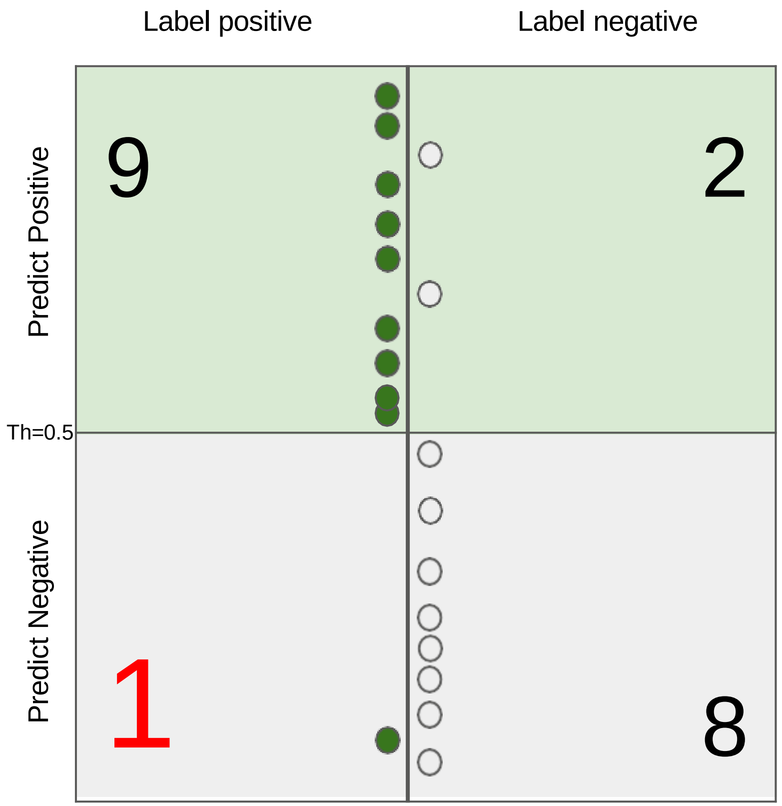
Th	TP	TN
0.5	9	8

Point metrics: False Positives



Th	TP	TN	FP
0.5	9	8	2

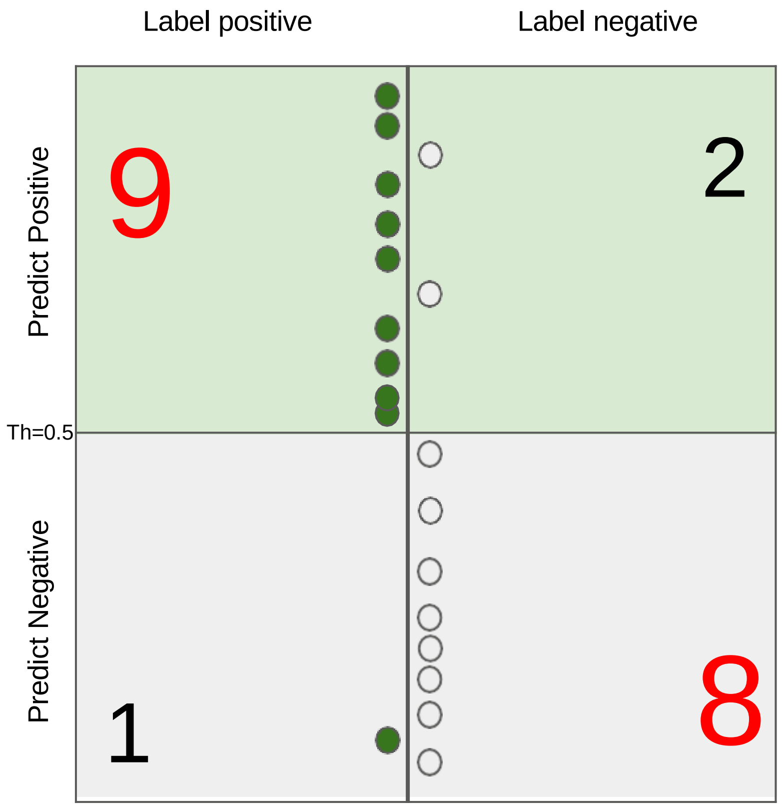
Point metrics: False Negatives



Th	TP	TN	FP	FN
0.5	9	8	2	1

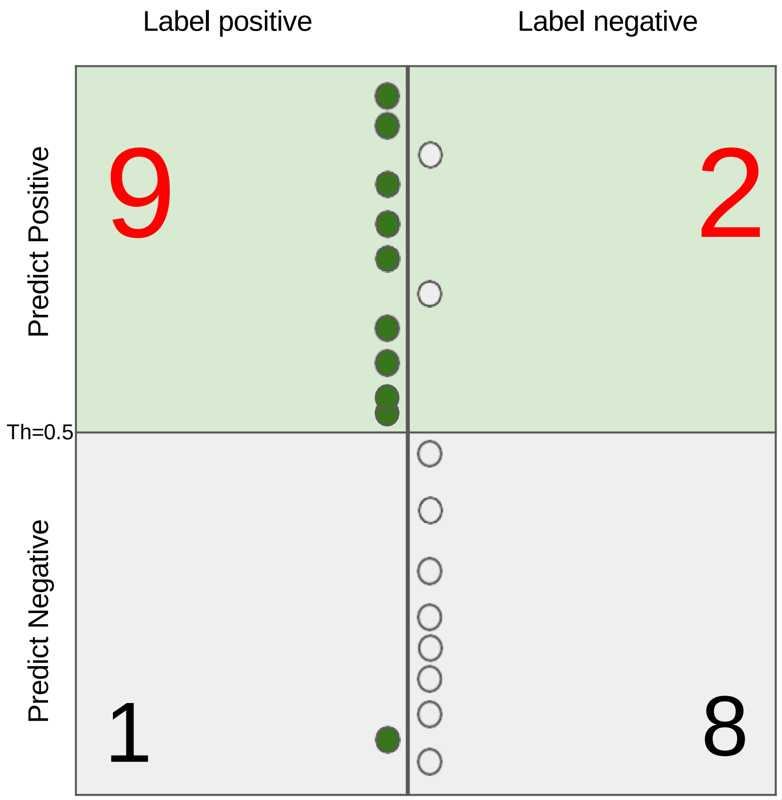
FP and FN also called Type-1 and Type-2 errors

Point metrics: Accuracy



Th	TP	TN	FP	FN	Acc
0.5	9	8	2	1	.85

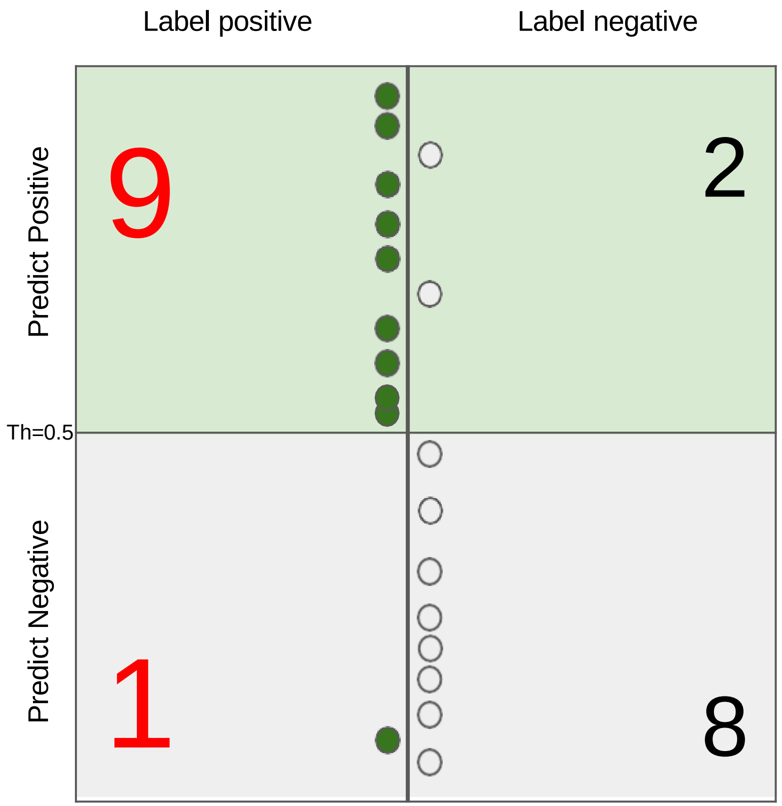
Point metrics: Precision



Th	TP	TN	FP	FN	Acc	Pr
0.5	9	8	2	1	.85	.81

$$P = \frac{TP}{TP + FP}$$

Point metrics: Positive Recall (Sensitivity)



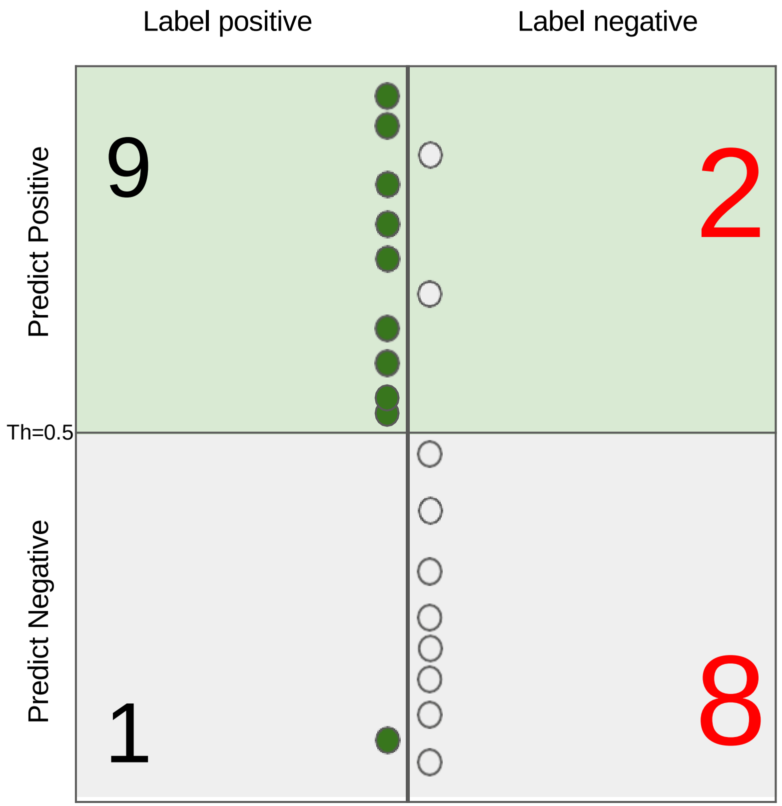
Th	TP	TN	FP	FN	Acc	Pr	Recall
0.5	9	8	2	1	.85	.81	.9

$$R = \frac{TP}{TP + FN}$$

Trivial 100% recall = pull everybody above the threshold.
Trivial 100% precision = push everybody below the threshold except 1 green on top.
(Hopefully no gray above it!)

Striving for good precision with 100% recall =
pulling up the lowest green as high as possible in the ranking.
Striving for good recall with 100% precision =
pushing down the top gray as low as possible in the ranking.

Point metrics: Negative Recall (Specificity)



Th	TP	TN	FP	FN	Acc	Pr	Recall	Spec
0.5	9	8	2	1	.85	.81	.9	0.8

$$Specificity = \frac{TN}{TN + FP}$$

Point metrics: F1-score

Th	TP	TN	FP	FN	Acc	Pr	Recall	Spec	F1
0.5	9	8	2	1	.85	.81	.9	.8	.857

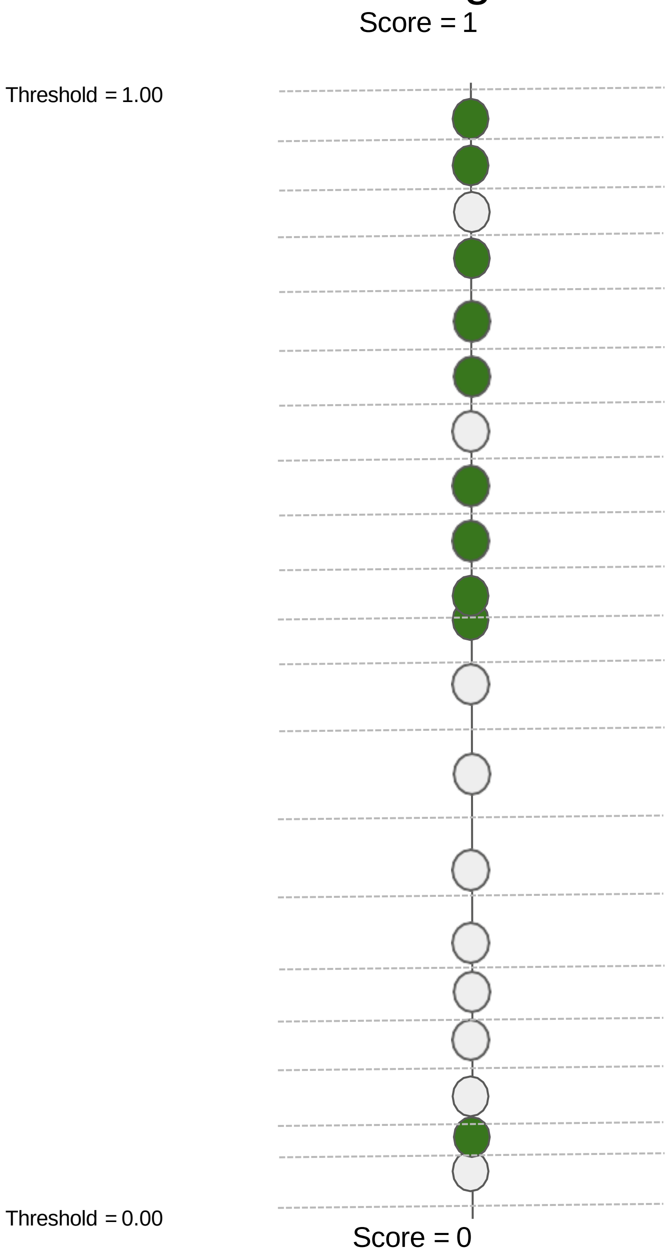
$$F_1 = \left(\frac{2}{\mathrm{recall}^{-1} + \mathrm{precision}^{-1}}\right) = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}.$$

Point metrics: Changing threshold

Th	TP	TN	FP	FN	Acc	Pr	Recall	Spec	F1
0.6	7	8	2	3	.75	.77	.7	.8	.733

effective thresholds = # examples + 1

Threshold Scanning



Threshold	TP	TN	FP	FN	Accuracy	Precision	Recall	Specificity	F1
1.00	0	10	0	10	0.50	1	0	1	0
0.95	1	10	0	9	0.55	1	0.1	1	0.182
0.90	2	10	0	8	0.60	1	0.2	1	0.333
0.85	2	9	1	8	0.55	0.667	0.2	0.9	0.308
0.80	3	9	1	7	0.60	0.750	0.3	0.9	0.429
0.75	4	9	1	6	0.65	0.800	0.4	0.9	0.533
0.70	5	9	1	5	0.70	0.833	0.5	0.9	0.625
0.65	5	8	2	5	0.65	0.714	0.5	0.8	0.588
0.60	6	8	2	4	0.70	0.750	0.6	0.8	0.667
0.55	7	8	2	3	0.75	0.778	0.7	0.8	0.737
0.50	8	8	2	2	0.80	0.800	0.8	0.8	0.800
0.45	9	8	2	1	0.85	0.818	0.9	0.8	0.857
0.40	9	7	3	1	0.80	0.750	0.9	0.7	0.818
0.35	9	6	4	1	0.75	0.692	0.9	0.6	0.783
0.30	9	5	5	1	0.70	0.643	0.9	0.5	0.750
0.25	9	4	6	1	0.65	0.600	0.9	0.4	0.720
0.20	9	3	7	1	0.60	0.562	0.9	0.3	0.692
0.15	9	2	8	1	0.55	0.529	0.9	0.2	0.667
0.10	9	1	9	1	0.50	0.500	0.9	0.1	0.643
0.05	10	1	9	0	0.55	0.526	1	0.1	0.690
0.00	10	0	10	0	0.50	0.500	1	0	0.667