Machine Learning

Evaluation Metrics 3

Mostafa S. Ibrahim
Teaching, Training and Coaching for more than a decade!

Artificial Intelligence & Computer Vision Researcher
PhD from Simon Fraser University - Canada
Bachelor / MSc from Cairo University - Egypt
Ex-(Software Engineer / ICPC World Finalist)

© 2023 All rights reserved.
Please do not reproduce or redistribute this work without permission from the author

Precision Metric

- A metric that measures the **accuracy** of **positive predictions**
  - How **reliable** are the positive predictions of the model?
  - To what degree the model minimizes the number of false positives
  - Useful when **false positives** have significant consequences, such as medical diagnostics or fraud detection (reject transaction of a good customer)
    - Observe FP in the dominator
- Formula
  - What is the model positive predictions? TP and FP
  - Which are the right positive predictions? TP
- Find a trivial way to get a 100% precision model?

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

- Return a single correct prediction
- Then TP = 1 and FP = 0
- Hence 1/1 = 100% precision!

Recall Metric

- Indicates how many of the true positive instances were correctly predicted
  - Goal: Find **as many positive** examples as possible (more tp out of p)
    - focuses **solely** on the **positive** instances
    - p = all gt positive = tp + fn. Hence with more tp, we get lower fn
  - A **higher recall** value indicates that the model has a **lower** rate of **false negatives**
    - Important when the cost of false negatives such in breast cancer
    - Observe FN in the dominator
      - **Constrained** with FN = P - TP
- Also known as True positive rate (TPR) / Sensitivity
- Find a trivial way to get a 100% recall model?

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

- Classify every example as positive, hence you found all P examples
- Then TP = P, FN = 0
- Hence P/P = 100% recall!

Precision vs Recall: Goal

- Precision focuses on the **accuracy** of **positive predictions** and avoid **false positives**
  - Use whenever the cost of false positive is higher than false negative
- Recall focuses on the model's ability to capture **all positive** instances and avoid **false negatives**
  - P = TP + FP
  - Use whenever the cost of false negative is higher than false positive

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

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

Precision vs Recall: The threshold

- To convert the probabilities to classifications (P/N), we use a threshold
- What happens when we use a high threshold like 0.9 compared to 0.5?
- A higher threshold means the model is very restrictive on predicting positive
  - As a result both tp and fp are *probably* reduced
  - As a result, from the equation, we *probably* get higher precision
- The opposite is true with lower threshold (higher tp/fp and lower precision)
  - When we get more and more positive labels (tp/fp), we get most of the tp
  - Hence, a higher recall
- Precision and Recall are typically inversely proportional to each other
  - But not necessary with very accurate models
- Achieving the desired **balance** between FP and TP depends on the specific **problem** and the **cost** of false positives and false negatives
  - With a **strong model**, you can get high precision and recall

	TP	FN P-TP	TN	FP N-TN	Precision	Recall	Accuracy	FPR
1	160	40	100	50	76	80	74	0.33
2	160	40	120	30	84	80	80	0.2
3	160	40	140	10	94	80	86	0.07
4	160	40	150	0	100	80	89	0.00
5	120	80	100	50	71	60	63	0.33
6	140	60	100	50	74	70	69	0.33
7	160	40	100	50	76	80	74	0.33
8	200	0	100	50	80	100	86	0.33
9	200	0	150	0	100	100	100	0
10	1	199	150	0	100	1	43	0
11	200	0	0	150	75	100	57	1

- Assume P = 200 positive examples
- And N = 150 negative examples

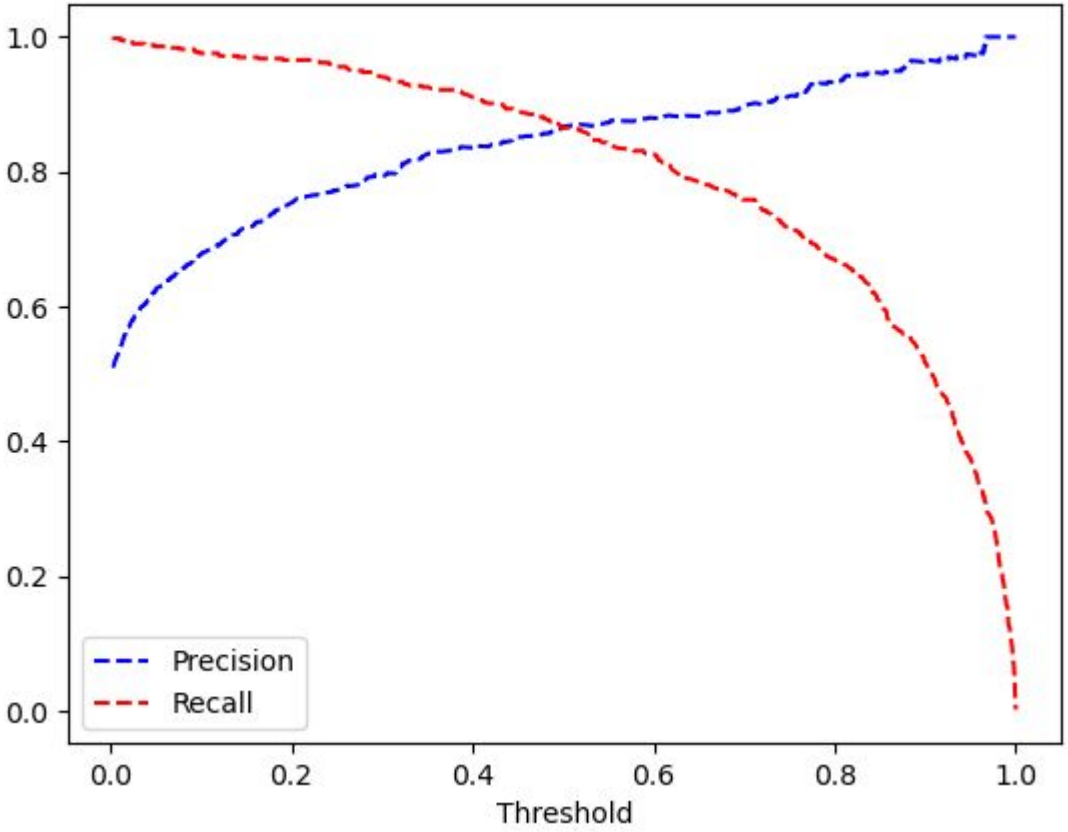
$$\frac{TP}{TP + FP}$$

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

Precision vs Recall: The threshold

- We can enumerate all the threshold, compute precision and recall and plot them
- The 2 curves will give us pretty good understanding for the behaviour
- Also visually, we can select a good threshold

Precision vs Recall: The threshold



For each threshold:

- Compute its precision
- Compute its recall

Draw the 2 curves

Precision vs Recall: The threshold

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import precision_recall_curve
from sklearn.datasets import make_classification

# create random binary classification task
X, y = make_classification(n_samples=5000, n_classes=2, random_state=42)
X_train, X_val, y_train, y_val = train_test_split(X, y, test_size=0.2, random_state=42)

model = LogisticRegression(random_state=42)
model.fit(X_train, y_train)
y_prop = model.predict_proba(X_val)[:, 1]   # 2nd column: prob class 1
```

Precision vs Recall: The threshold

precision_recall_curve:
Compute precision-recall **pairs** for different probability **thresholds**

```
# The last precision and recall values are 1. and 0. respectively
# and do not have a corresponding threshold
precision, recall, threshold = precision_recall_curve(y_val, y_prop)
# Remove last 2 values
precision, recall = precision[:-1], recall[:-1]

plt.plot(threshold, precision, 'b--', label='Precision')
plt.plot(threshold, recall, 'r--', label='Recall')
plt.xlabel('Threshold')
plt.legend(loc='lower left')
```

Target Precision or Recall Performance

- Given a **specific precision** (or recall), we can find the optimal threshold

```
35
36
37  # precision is increasing
38  target_precision = 0.935
39  best_threshold_idx = np.argmax(precision >= target_precision)
40  best_threshold = threshold[best_threshold_idx]
41  print(f'Threshold {best_threshold} has precision >= {target_precision} '
42        f'has recall {recall[best_threshold_idx]}')
43  # Threshold 0.8002828287526059 has precision >= 0.935 has recall 0.670020120724346
44
45  # recall is decreasing
46  target_recall = 0.82
47  best_threshold_idx = np.argmin(recall >= target_recall)
48  best_threshold = threshold[best_threshold_idx]
49  print(f'Threshold {best_threshold} has recall >= {target_recall} '
50        f'has precision {precision[best_threshold_idx]}')
51
52  # Threshold 0.604938756634422 has recall >= 0.82 has precision 0.8790496760259179
53
```

"Acquire knowledge and impart it to the people."

"Seek knowledge from the Cradle to the Grave."