

pr auc, roc auc 的区别和使用区别，请举例说明

以下是关于 PR AUC (Precision-Recall AUC) 和 ROC AUC (Area Under the ROC Curve) 的详细对比和实际使用区别，结合具体例子说明：

1. 核心区别总结

指标 PR AUC ROC AUC

横纵轴 横轴：Recall (召回率)

纵轴：Precision (精确率) 横轴：FPR (假正率)

纵轴：TPR (召回率)

关注点 正样本的预测质量 (少数类) 整体分类能力 (正负类平衡时)

数据不平衡影响 敏感 (正样本极少时更实用) 不敏感 (正负样本比例变化时较稳定)

最佳值 1 (完美分类) 1 (完美分类)

最差值 正样本比例 (基线) 0.5 (随机猜测)

2. 使用场景对比

何时用 PR AUC?

正样本极少 (如点击率预测、疾病检测、欺诈检测)。

更关注正样本的准确性 (如广告推荐中避免误推低质量广告)。

负样本数量远大于正样本 (如100:1的不平衡数据)。

何时用 ROC AUC?

正负样本相对平衡 (如1:1到1:10)。

需要全面评估模型整体性能 (如信用评分、垃圾邮件分类)。

比较不同模型的普适性能。

3. 实际例子说明

案例1: 广告点击预测 (正样本1%)

数据: 10,000次曝光, 100次点击 (正样本), 9,900次未点击 (负样本)。

模型A:

预测90次点击 (其中80次正确, 10次错误)。

PR AUC计算:

$$\text{Precision} = 80/90 \approx 0.89$$

$$\text{Recall} = 80/100 = 0.8$$

PR曲线下面积反映模型对正样本的捕捉能力。

ROC AUC计算：

$$\text{TPR} = 80/100 = 0.8$$

$$\text{FPR} = 10/9,900 \approx 0.001$$

ROC曲线下面积反映整体区分能力。

结果：

PR AUC=0.85，ROC AUC=0.95。

解读：虽然ROC AUC高，但PR AUC更能反映模型对稀疏点击的预测质量。

案例2：癌症检测（正样本5%）

数据：1,000例，50例癌症（正），950例健康（负）。

模型B：

预测40例癌症（其中30例正确，10例误诊）。

PR AUC：

$$\text{Precision} = 30/40 = 0.75$$

$$\text{Recall} = 30/50 = 0.6$$

ROC AUC：

$$\text{TPR} = 0.6, \text{FPR} = 10/950 \approx 0.01$$

结果：

PR AUC=0.7，ROC AUC=0.92。

解读：ROC AUC虚高（因负样本多），PR AUC更真实反映漏诊风险。

4. 曲线形状对比

PR曲线

基线：水平线

=

正样本比例

$y = \text{正样本比例}$ （如正样本1%时基线=0.01）。

好模型：曲线向右上凸起，接近(1,1)。

差模型：曲线靠近基线。

ROC曲线

基线：对角线

=

$y = x$ （随机猜测）。

好模型：曲线向左上凸起，接近(0,1)。

差模型：靠近对角线。

https://miro.medium.com/max/1400/1*4QSFqYkg0dQfij5h0kKv5A.png

5. 代码示例 (Python)

python

```
from sklearn.metrics import precision_recall_curve, roc_curve, auc
```

```
# 假设真实标签和预测概率
```

```
y_true = [0, 1, 0, 0, 1, 0, 1, 0, 0, 1] # 正样本占40%
```

```
y_scores = [0.1, 0.9, 0.2, 0.3, 0.8, 0.05, 0.7, 0.1, 0.4, 0.6]
```

```
# 计算PR AUC
```

```
precision, recall, _ = precision_recall_curve(y_true, y_scores)
```

```
pr_auc = auc(recall, precision) # 注意参数顺序是recall在前！
```

```
# 计算ROC AUC
```

```
fpr, tpr, _ = roc_curve(y_true, y_scores)
```

```
roc_auc = auc(fpr, tpr)
```

```
print(f"PR AUC: {pr_auc:.3f}, ROC AUC: {roc_auc:.3f}")
```

输出：

```
text
```

```
PR AUC: 0.917, ROC AUC: 0.938
```

6. 如何选择？

优先PR AUC：

正样本<10%

业务更关注精确率（如推荐系统要求精准推送）

优先ROC AUC：

正负样本较平衡

需要评估整体分类能力（如人脸识别）

7. 经典论文结论

Davis & Goadrich (2006) 指出：在高度不平衡数据中，PR曲线比ROC曲线更能揭示模型性能差异。

"When dealing with highly skewed datasets, PR curves give a more informative picture of an algorithm's performance."

在二元分類問題中，**PR AUC (Precision-Recall AUC)** 和 **ROC AUC (Receiver Operating Characteristic AUC)** 都是評估模型性能的重要指標，尤其是在處理資料不平衡問題時。雖然兩者都衡量了模型區分正負類別的能力，但它們關注的角度和適用場景有所不同。

PR AUC vs. ROC AUC：核心區別

ROC 曲線 (Receiver Operating Characteristic Curve) 和 ROC AUC

- 定義：ROC 曲線以 真陽性率 (True Positive Rate, TPR) 作為 Y 軸，假陽性率 (False Positive Rate, FPR) 作為 X 軸繪製。
 - **TPR (召回率/Recall/敏感度)**: $TP/(TP+FN)$ - 所有實際為正的樣本中，被正確預測為正的比例。
 - **FPR**: $FP/(FP+TN)$ - 所有實際為負的樣本中，被錯誤預測為正的比例。
- 如何繪製：透過不斷調整分類閾值 (threshold)，計算在每個閾值下的 TPR 和 FPR，然後將這些點連接起來形成曲線。
- **ROC AUC 的意義**：曲線下面積 (Area Under the Curve)。AUC 的值介於 0 到 1 之間，越接近 1 表示模型的分類性能越好。它代表了模型隨機選擇一個正樣本和一個負樣本時，正樣本的預測得分高於負樣本的機率。
- 特點：ROC 曲線對類別分佈不敏感。即使正負樣本的比例發生很大變化，ROC 曲線的形狀和 ROC AUC 的值也不太會受到影響。這是因為 FPR 和 TPR 的計算都分別基於各自類別的總數。

PR 曲線 (Precision-Recall Curve) 和 PR AUC

- 定義：PR 曲線以 精確率 (**Precision**) 作為 Y 軸，召回率 (**Recall**) 作為 X 軸繪製。
 - **Precision** (精確率): $TP/(TP+FP)$ - 所有被模型預測為正的樣本中，真正是正樣本的比例。
 - **Recall** (召回率): $TP/(TP+FN)$ - 同上，真陽性率。
- 如何繪製：同樣透過調整分類閾值，計算在每個閾值下的 Precision 和 Recall，然後將這些點連接起來形成曲線。
- **PR AUC** 的意義：曲線下面積。AUC 的值介於 0 到 1 之間，越接近 1 表示模型性能越好。它代表了在召回率從 0 到 1 的範圍內，精確率的平均值。
- 特點：PR 曲線對類別分佈敏感。它更關注少數類別（正樣本）的性能，特別是在資料集高度不平衡時，PR 曲線能夠更真實地反映模型的表現。

使用場景區別與範例

理解這兩者的核心區別，在選擇正確的評估指標時至關重要：

何時使用 ROC AUC?

- 資料集相對平衡，或者正負類都同等重要時。
- 當你關心的是模型區分兩類的能力，而不論類別分佈如何。
- 當你希望比較不同模型在廣泛的分類閾值下的表現時。
- 範例：疾病診斷。如果誤診（假陽性）和漏診（假陰性）的成本都比較高且相對平衡，那麼 ROC AUC 是一個不錯的選擇。你希望模型能夠很好地將患病者和健康者區分開來。

何時使用 **PR AUC**?

- 資料集高度不平衡時，特別是當你對少數類別（正類）的預測性能更感興趣時。
- 當你更關心精確率 (Precision)和召回率 (Recall)**之間的權衡時。例如，在廣告點擊預測中，你可能希望在確保足夠召回率的同時，維持較高的精確率，以避免對大量不點擊的用戶投放廣告。
- 當正類是稀有事件時（例如，詐騙檢測、點擊預測、罕見疾病）。

2. Key Differences

Aspect	ROC AUC	PR AUC	🔗
X-Axis	False Positive Rate (FPR)	Recall (TPR)	
Y-Axis	True Positive Rate (TPR, Recall)	Precision	
Focus	Balances positive and negative class performance	Focuses on positive class performance	
Class Imbalance	Robust, considers negatives (FPR)	Sensitive, ignores true negatives	
Interpretation	Ability to distinguish classes	Trade-off between precision and recall	
Use Case	Balanced datasets or when negatives matter	Imbalanced datasets, positive class is critical	
Sensitivity	Less sensitive to rare positives	Highly sensitive to rare positives	

3. Use Cases and When to Use

ROC AUC

- **When to Use:**
 - **Balanced Datasets:** When positive and negative classes are roughly equal in size.
 - **Equal Importance of Classes:** When both false positives and false negatives have similar costs.
 - **General Performance:** When you want to evaluate the model's overall ability to distinguish classes.

PR AUC

- **When to Use:**
 - **Imbalanced Datasets:** When the positive class is rare (e.g., 1% of users purchase tickets).
 - **Positive Class Critical:** When precision (avoiding false positives) or recall (capturing all positives) is more important.
 - **Recommendation Systems:** When recommending relevant items (e.g., events) where false positives reduce user trust.
-

Reference:

1. ROC-AUC 与 PR-AUC 的区别与联系

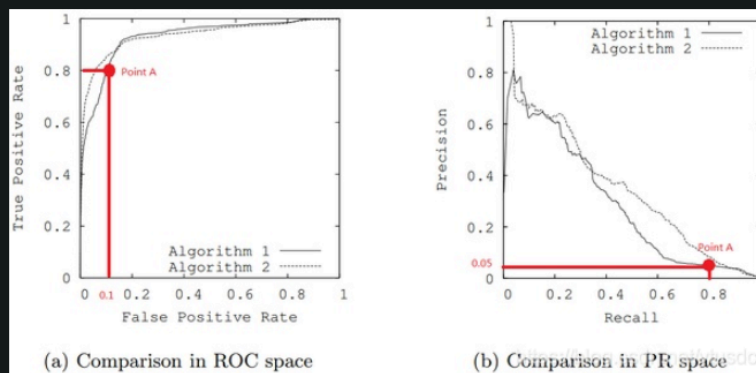
<https://zhuanlan.zhihu.com/p/349516115>

2.

<https://blog.csdn.net/ytusdc/article/details/107699160>

PRC曲线的优势

在正负样本分布得极不均匀(**highly skewed datasets**), 负例远大于正例时, 并且这正是该问题正常的样本分布时, **PRC比ROC能更有效地反应分类器的好坏, 即PRC曲线在正负样本比例悬殊较大时更能反映分类的真实性能**。例如上面的(c)(d)中正负样本比例为1:10, ROC效果依然看似很好, 但是PR曲线则表现的比较差。举个例子,



单从图 (a) 看, 这两个分类器都比较完美(非常接近左上角)。而从图 (b) 可以看出, 这两个分类器仍有巨大的提升空间。那么原因是什么呢? 通过看Algorithm1的点 A, 可以得出一些结论。首先图 (a) 和 (b) 中的点A是相同的点, 因为TPR就是Recall, 两者是一样的。

单从图（a）看，这两个分类器都比较完美(非常接近左上角)。而从图（b）可以看出，这两个分类器仍有巨大的提升空间。那么原因是什么呢？通过看Algorithm1的点A，可以得出一些结论。首先图（a）和（b中）的点A是相同的点，因为TPR就是Recall，两者是一样的。

假设数据集有100个正样本。可以得到以下结论(反推样本个数)：

由图（a）点A，可得： $TPR = TP / (TP + FN) = TP / \text{所有正样本} = TP / 100 = 0.8$ ，所以 $TP = 80$ 。

由图（b）点A，可得： $Precision = TP / (TP + FP) = 80 / (80 + FP) = 0.05$ ，所以 $FP = 1520$ 。

再由图（a）点A，可得： $FPR = FP / (FP + TN) = FP / \text{所有负样本} = 1520 / \text{所有负样本} = 0.1$ ，所以负样本数量是15200。

由此，可以得出原数据集中只有100个正样本，却有15200个负样本！这就是极不均匀的数据集。直观地说，在点A处，分类器将1600 (1520+80)个样本预测为positive，而其中实际上只有80个是真正的positive。我们凭直觉来看，其实这个分类器并不好。但**由于真正negative instances的数量远远大于positive，ROC的结果却“看上去很美”**，因为这时FPR因为负例基数大的缘故依然很小。所以，在这种情况下，PRC更能体现本质。

ROC曲线与PRC曲线表现差异的原因

为什么会有上面分析到的两者差异呢？下面摘自引用[1]的解释很清楚，FPR 和 TPR (Recall) 只与真实的正例或负例中的一个相关（可以从他们的计算公式中看到），而其他指标如Precision则同时与真实的正例与负例都有关，即下面文字说的“both columns”，这可以结合混淆矩阵和各指标的计算公式明显看到。

觉得还不错？ 一键收藏

ROC曲线与PRC曲线表现差异的原因

为什么会有上面分析到的两者差异呢？下面摘自引用[1]的解释很清楚，FPR 和 TPR (Recall) 只与真实的正例或负例中的一个相关（可以从他们的计算公式中看到），而其他指标如Precision则同时与真实的正例与负例都有关，即下面文字说的“both columns”，这可以结合混淆矩阵和各指标的计算公式明显看到。

ROC曲线 vs PR曲线实际应用

****这有什么实际意义呢？****在很多实际问题中，正负样本数量往往很不均衡。比如，计算广告领域经常涉及转化率模型，正样本的数量往往是负样本数量的1/1000甚至1/10000。若选择不同的测试集，P-R曲线的变化就会非常大，而ROC曲线则能够更加稳定地反映模型本身的好坏。所以，ROC曲线的适用场景更多，被广泛用于排序、推荐、广告等领域。

需要注意的是，选择**P-R曲线**还是**ROC曲线**是因实际问题而异的，如果研究者希望更多地看到模型在特定数据集上的表现，**P-R曲线**则能够更直观地反映其性能。”

结论--ROC vs PR 各自使用场景

- 1.调优观察模型时，使用auc值会更好衡量模型性能，因为受到测试集和验证集的分布影响小。
- 2.模型调优完毕，还需准备一份需求可能出现的分布的测试集，将这份测试集输入模型，输出一个精确率和召回率，作为给业务方模型的预期效果，也可以直观看出模型对现实可能出现的数据集的效果