

# 问题

为什么使用F1 score？（这里主要讨论为何使用 F1 score 而不是算术平均）

## F1 score

F1 score是分类问题中常用的评价指标，定义为精确率（Precision）和召回率（Recall）的调和平均数。

$$F1 = \frac{1}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (1)$$

补充一下精确率和召回率的公式：

TP（True Positive）：真正例

FP（False Positive）：假正例

FN（False Negative）：假反例

TN（True Negative）：真反例

**精确率（Precision）**：  $Precision = \frac{TP}{TP+FP}$

**召回率（Recall）**：  $Recall = \frac{TP}{TP+FN}$

精确率，也称为查准率，衡量的是预测结果为正例的样本中被正确分类的正例样本的比例。

召回率，也称为查全率，衡量的是真实情况下的所有正样本中被正确分类的正样本的比例。

F1 score 综合考虑了精确率和召回率，其结果更偏向于 Precision 和 Recall 中较小的那个，即 Precision 和 Recall 中较小的那个对 F1 score 的结果取决定性作用。例如若  $Precision = 1, Recall \approx 0$ ，由F1 score的计算公式可以看出，此时其结果主要受 Recall 影响。

如果对 Precision 和 Recall 取算术平均值（ $\frac{Precision+Recall}{2}$ ），对于  $Precision = 1, Recall \approx 0$ ，其结果约为 0.5，而 F1 score 调和平均的结果约为 0。这也是为什么很多应用场景中会选择使用 **F1 score** 调和平均值而不是算术平均值的原因，因为我们希望这个结果可以更好地反映模型的性能好坏，而不是直接平均模糊化了 **Precision** 和 **Recall** 各自对模型的影响。

补充另外两种评价方法：

**加权调和平均：**

上面的 F1 score 中，Precision 和 Recall 是同等重要的，而有的时候可能希望我们的模型更关注其中的某一个指标，这时可以使用加权调和平均：

$$F_{\beta} = (1 + \beta^2) \frac{1}{\frac{1}{Precision} + \beta^2 \times \frac{1}{Recall}} = (1 + \beta^2) \frac{Precision \times Recall}{\beta^2 \times Precision + Recall} \quad (2)$$

当  $\beta > 1$  时召回率有更大影响， $\beta < 1$  时精确率有更大影响， $\beta = 1$  时退化为 F1 score。

几何平均数：

$$G = \sqrt{Precision \times Recall} \quad (3)$$

## 参考资料

---

为什么要用f1-score而不是平均值 <https://www.cnblogs.com/walter-xh/p/11140715.html>