

## 1.异常检测二分类中通常将异常点设置为 0 异常点设置为 1 的分



在二分类异常检测场景中，标签向量定义正常点是使用1还是0更有助于提升二分类异常检测的性能？

Thought for 7 seconds ▾

在二分类异常检测场景中，通常将异常点（异常样本）的标签设为 1，\*\*正常点（正常样本）的标签设为 0，这样更有助于提升异常检测的性能。

### 原因分析

#### 1. 统一的惯例和标准

- **正类 (Positive Class)** 表示感兴趣的目标类别：在二分类任务中，标签为 1 的类别通常被视为正类，即模型要重点识别或检测的目标。
- **异常检测的目标是识别异常样本**：因此，将异常样本标记为 1，符合模型训练和评估的标准惯例。

#### 2. 评估指标的计算方式

- **指标针对正类计算**：诸如精确率 (Precision)、召回率 (Recall)、F1 值等评估指标，都是针对标签为 1 的正类进行计算。
  - **精确率**：预测为异常的样本中，实际为异常的比例。
  - **召回率**：实际异常样本中，被正确识别为异常的比例。
- **将异常样本设为正类，可以直接使用这些指标评估模型对异常检测的性能。**

#### 3. 损失函数和模型优化

- **某些损失函数对正负类处理不同**：例如，在不平衡数据中，可能对正类（异常样本）给予更大的权重，以解决类别不平衡问题。
- **更好的梯度信息**：将异常样本设为正类，可以使模型在训练过程中更关注异常样本，提高模型对异常的敏感度。

#### 4. 易于处理类别不平衡

- **类别不平衡是异常检测的常见问题**：正常样本数量远多于异常样本。
- **通过调整正类权重**：可以在损失函数中增加正类的权重，促使模型更重视异常样本的识别。