auc计算方式——python实现

Lucifer 关注他

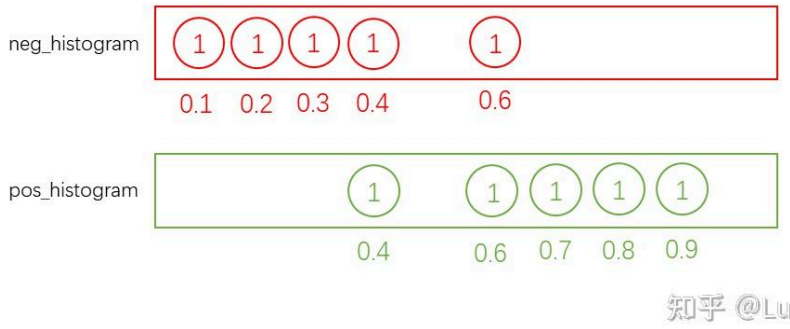
关于auc的定义和原理，我就不在赘述了，已经有比较多的文章说过了。

我们来一起看看auc的python实现方式，尤其是正负例相等的情况下。（有些文章并没有说明这种情况）

```
import numpy as np


def auc_calculate(labels, preds, n_bins=100):
    postive_len = sum(labels)  # 正样本数量（因为正样本都是1）
    negative_len = len(labels) - postive_len  # 负样本数量
    total_case = postive_len * negative_len  # 正负样本对
    pos_histogram = [0 for _ in range(n_bins)]
    neg_histogram = [0 for _ in range(n_bins)]
    bin_width = 1.0 / n_bins
    for i in range(len(labels)):
        nth_bin = int(preds[i] / bin_width)
        if labels[i] == 1:
            pos_histogram[nth_bin] += 1
        else:
            neg_histogram[nth_bin] += 1
    accumulated_neg = 0
    satisfied_pair = 0
    for i in range(n_bins):
        satisfied_pair += (pos_histogram[i] * accumulated_neg + pos_histogram[i] * neg
        accumulated_neg += neg_histogram[i]
        # print(i,satisfied_pair,accumulated_neg)
    return satisfied_pair / float(total_case)


if __name__ == '__main__':
    y = np.array([1, 1, 1, 1, 0, 1, 0, 0, 0, 0])
    pred = np.array([0.9, 0.8, 0.7, 0.6, 0.6, 0.4, 0.4, 0.3, 0.2, 0.1])

    print("auc:", auc_calculate(y, pred))
```

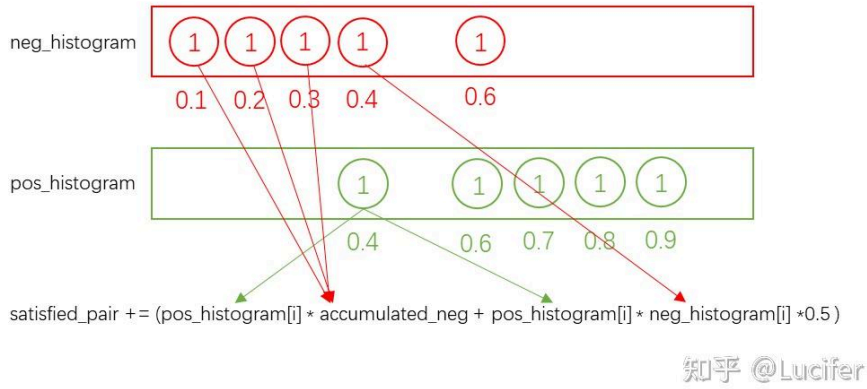
对照上述代码，我们用图解的方式看一下过程：

首先生成neg_histogram和pos_histogram，用于存放样本



neg_

计算每个正样本前有几个负样本，**如果相等的话，记为0.5**（对应的roc曲线为一条斜线）。



satisfied_pair 计算公式

以0.4这个正样本为例，它在所有3个负样本前面，且有一个负样本和它相等，因此记为3.5

以此类推，最后的 satisfied_pair = 3.5 + 4.5 + 5 + 5 + 5

正负样本对 total_case = 5 * 5

所以auc = satisfied_pair / total_case = 23/25 = 0.92

让我们通过sklearn验证一下：

```
from sklearn import metrics

def aucfun(act, pred):
    fpr, tpr, thresholds = metrics.roc_curve(act, pred, pos_label=1)
    return metrics.auc(fpr, tpr)

if __name__ == '__main__':
    y = np.array([1, 1, 1, 1, 0, 1, 0, 0, 0, 0])
    pred = np.array([0.9, 0.8, 0.7, 0.6, 0.6, 0.4, 0.4, 0.3, 0.2, 0.1])

    print(aucfun(y, pred))
```

```
auc: 0.92
sklearn: 0.92
```
知乎 @Lucifer

结果是一致的，我们下次再见

发布于 2022-02-16 18:13

Python auc 机器学习

理性发言，友善互动

2 条评论 默认 最新

约瑟福
前提是预测分数已经排好序了？
2022-04-07 回复 喜欢

Luoran
新粉一枚！希望作者多多输出