

Python 3 玩儿转机器学习

讲师：liuyubobobo

版权所有 侵权必究
liuyubobobo

慕课网《Python3机器学习》

分类算法的评价

讲师：liuyuboboo

版权所有，侵权必究

慕课网《Python3机器学习》

分类准确度的问题

讲师：liuyuboboo

版权所有，侵权必究

分类准确度的问题

一个癌症预测系统，输入体检信息，可以判断是否有癌症

预测准确度：99.9%

是好？是坏？

分类准确度的问题

一个癌症预测系统，输入体检信息，可以判断是否有癌症

预测准确度：99.9%

如果癌症产生的概率只有0.1%

我们的系统预测所有人都是健康，即可达到99.9%的准确率

分类准确度的问题

一个癌症预测系统，输入体检信息，可以判断是否有癌症

预测准确度：99.9%

如果癌症产生的概率只有0.01%

我们的系统预测所有人都是健康，可达到99.99%的准确率

分类准确度的问题

对于极度偏斜(Skewed Data)的数据,
只使用分类准确度是远远不够的

使用混淆矩阵做进一步的分析

混淆矩阵 Confusion Matrix

对于二分类问题

	实际正类	实际负类
预测正类	真正类 (TP)	假正类 (FP)
预测负类	假负类 (FN)	真负类 (TN)

混淆矩阵 Confusion Matrix

对于二分类问题

行代表真实值

混淆矩阵 Confusion Matrix

对于二分类问题

行代表真实值

0		
1		

混淆矩阵 Confusion Matrix

对于二分类问题

行代表真实值

	0	1
0		
1		

列代表预测值

混淆矩阵 Confusion Matrix

对于二分类问题

行代表真实值

	0	1
0		
1		

列代表预测值

混淆矩阵 Confusion Matrix

对于二分类问题

行代表真实值

	0	1
0		
1		

列代表预测值

0 - Negative

1 - Positive

混淆矩阵 Confusion Matrix

对于二分类问题

行代表真实值

	0	1
0	预测negative正确	
1		

列代表预测值

0 - Negative

1 - Positive

混淆矩阵 Confusion Matrix

对于二分类问题

行代表真实值

	0	1
0	预测negative正确 TN	
1		

列代表预测值

0 - Negative

1 - Positive

混淆矩阵 Confusion Matrix

对于二分类问题

行代表真实值

	0	1
0	预测negative正确 TN	预测positive错误
1		

列代表预测值

0 - Negative

1 - Positive

混淆矩阵 Confusion Matrix

对于二分类问题

行代表真实值

	0	1
0	预测negative正确 TN	预测positive错误 FP
1		

列代表预测值

0 - Negative

1 - Positive

混淆矩阵 Confusion Matrix

对于二分类问题

行代表真实值

	0	1
0	预测negative正确 TN	预测positive错误 FP
1	预测negative错误	

列代表预测值

0 - Negative

1 - Positive

混淆矩阵 Confusion Matrix

对于二分类问题

行代表真实值

	0	1
0	预测negative正确 TN	预测positive错误 FP
1	预测negative错误 FN	

列代表预测值

0 - Negative

1 - Positive

混淆矩阵 Confusion Matrix

对于二分类问题

行代表真实值

	0	1
0	预测negative正确 TN	预测positive错误 FP
1	预测negative错误 FN	预测positive正确

列代表预测值

0 - Negative

1 - Positive

混淆矩阵 Confusion Matrix

对于二分类问题

行代表真实值

	0	1
0	预测negative正确 TN	预测positive错误 FP
1	预测negative错误 FN	预测positive正确 TP

列代表预测值

0 - Negative

1 - Positive

混淆矩阵 Confusion Matrix

有10000个人

真实 \ 预测	0	1
0		
1		

混淆矩阵 Confusion Matrix

有10000个人

真实 \ 预测	0	1
0	9978	12
1	2	8

慕课网《Python3机器学习》

精准率和召回率

讲师：liuyubobobo

版权所有，侵权必究

精准率和召回率

有10000个人

真实 \ 预测	0	1
0	9978	12
1	2	8

精准率和召回率

真实 \ 预测	0	1
0	9978	12
1	2	8

精准率

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

精准率和召回率

真实 \ 预测	0	1
0	9978 TN	12 FP
1	2 FN	8 TP

精准率

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

$$精准率 = 8 / (8 + 12) = 40\%$$

精准率和召回率

真实 \ 预测	0	1
0	9978 TN	12 FP
1	2 FN	8 TP

精准率

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

$$精准率 = 8 / (8 + 12) = 40\%$$

精准率和召回率

真实 \ 预测	0	1
0	9978 TN	12 FP
1	2 FN	8 TP

召回率

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

$$\text{召回率} = 8 / (8 + 2) = 80\%$$

精准率和召回率

真实 \ 预测	0	1
0	9978 TN	12 FP
1	2 FN	8 TP

召回率

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

$$召回率 = 8 / (8+2) = 80\%$$

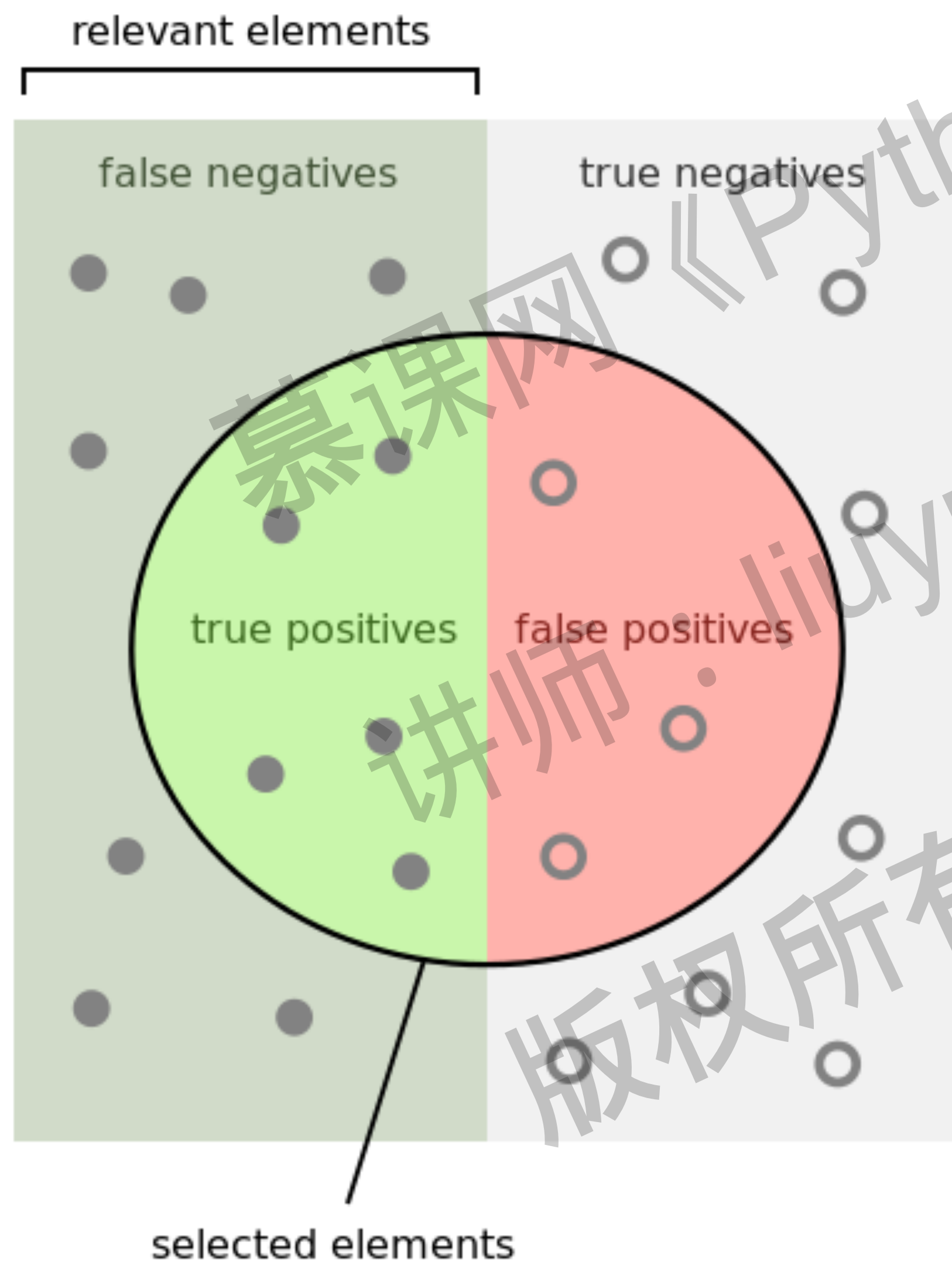
精准率和召回率

真实 \ 预测	0	1
0	9978 TN	12 FP
1	2 FN	8 TP

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

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

精准率和召回率



How many selected items are relevant?

Precision =



How many relevant items are selected?

Recall =



精准率和召回率

有10000个人，我们预测所有的人都是健康的

真实 \ 预测	0	1
0	9990	0
1	10	0

精准率和召回率

有10000个人，我们预测所有的人都是健康的

真实 \ 预测	0	1
0	9990	0
1	10	0

准确率 = 99.9%

精准率 = $0 / (0+0)$ 无意义

召回率 = $0 / (10+0) = 0$

慕课网《Python3机器学习》
实践：编写混淆矩阵，精准率和召回率

讲师：liuyubobobo
版权所有，侵权必究

慕课网《Python3机器学习》

F1 Score

讲师：liuyubobobo

版权所有，侵权必究

精准率和召回率

真实 \ 预测	0	1
0	9978 TN	12 FP
1	2 FN	8 TP

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

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

精准率和召回率

有的时候我们注重精准率。如股票预测

	0	1
0	预测negative正确 TN	预测posotive错误 FP
1	预测negative错误 FN	预测positive正确 TP

精准率

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

精准率和召回率

有的时候我们注重召回率。如病人诊断

	0	1
0	预测negative正确 TN	预测posotive错误 FP
1	预测negative错误 FN	预测positive正确 TP

召回率

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

F1 Score

二者都兼顾：F1 Score

$$F1 = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

F1 Score

F1 Score 是 precision 和 recall 的调和平均值

$$\frac{1}{F1} = \frac{1}{2} \left(\frac{1}{precision} + \frac{1}{recall} \right)$$

F1 Score

$$\frac{1}{F1} = \frac{1}{2} \left(\frac{1}{precision} + \frac{1}{recall} \right)$$

$$\frac{1}{F1} = \frac{1}{2} \left(\frac{precision + recall}{precision \cdot recall} \right)$$

$$F1 = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

慕课网《Python3机器学习》

实践：F1 Score

讲师：liuyubobobo

版权所有，侵权必究

慕课网《Python3机器学习》

Precision-Recall 的平衡

讲师：liuyubobobo

版权所有，侵权必究

逻辑回归 Logistic Regression

$$\hat{p} = \sigma(\theta^T \cdot x_b) = \frac{1}{1 + e^{-\theta^T \cdot x_b}}$$

$$\hat{y} = \begin{cases} 1, & \hat{p} \geq 0.5 & \theta^T \cdot x_b \geq 0 \\ 0, & \hat{p} < 0.5 & \theta^T \cdot x_b < 0 \end{cases}$$

决策边界

$$\theta^T \cdot x_b = 0$$

逻辑回归 Logistic Regression

决策边界

$$\theta^T \cdot x_b = 0$$

逻辑回归 Logistic Regression

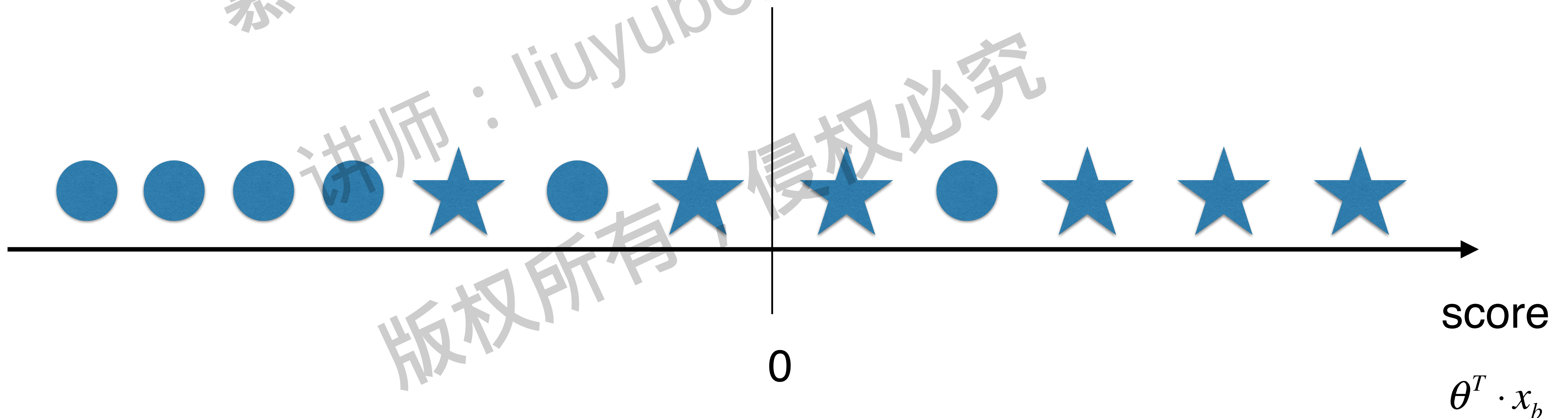
决策边界

$$\theta^T \cdot x_b = threshold$$

Precision-Recall 的平衡

精准率: $4/5=0.80$

召回率: $4/6=0.67$



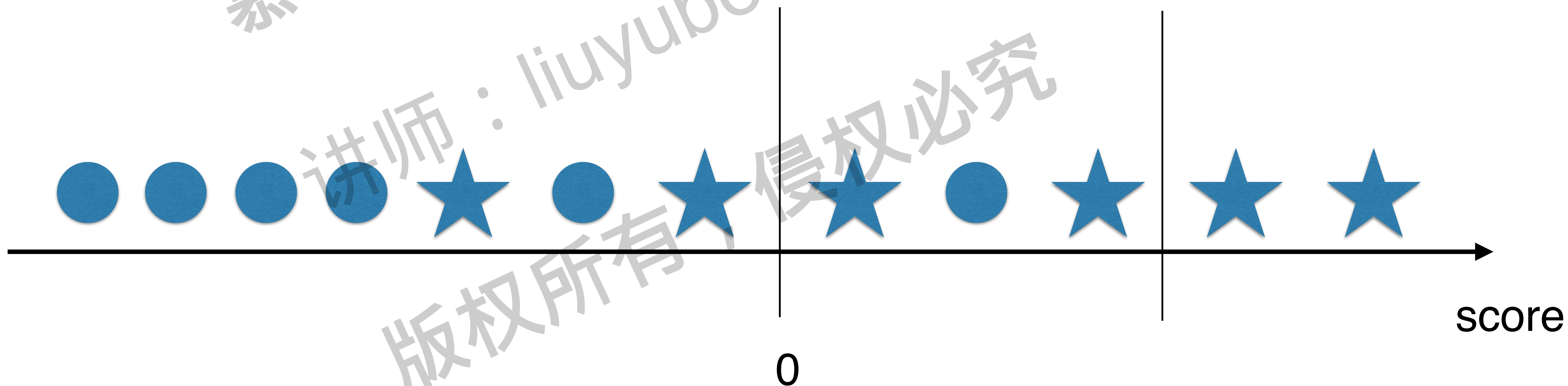
Precision-Recall 的平衡

精准率: $4/5=0.80$

精准率: $2/2=1.00$

召回率: $4/6=0.67$

召回率: $2/6=0.33$



Precision-Recall 的平衡

精准率: $6/8=0.75$

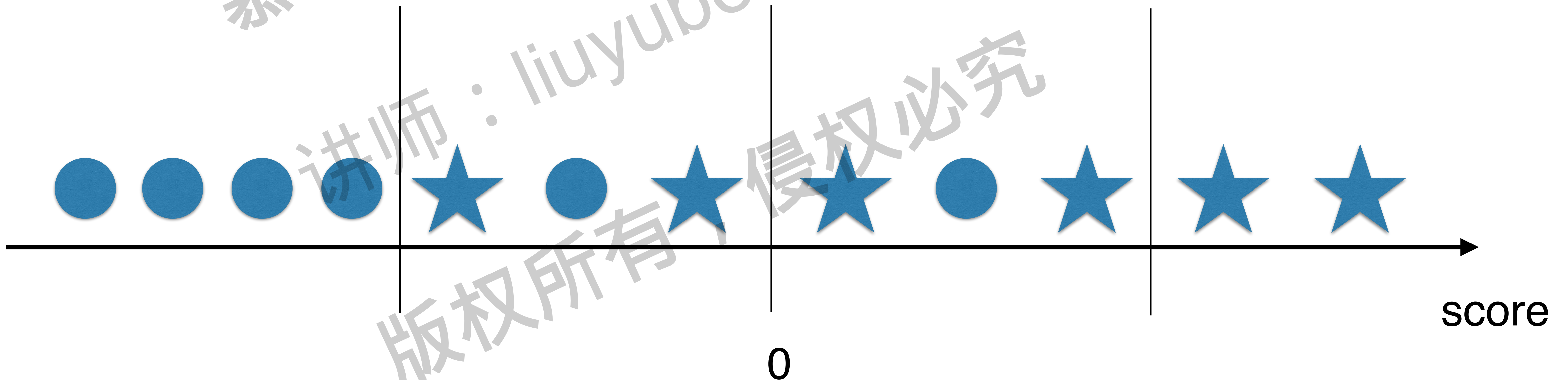
精准率: $4/5=0.80$

精准率: $2/2=1.00$

召回率: $6/6=1.00$

召回率: $4/6=0.67$

召回率: $2/6=0.33$



实践：scikit-learn中调节threshold

讲师：liuyubobobo
版权所有，侵权必究

慕课网《Python3机器学习》

Precision-Recall 曲线

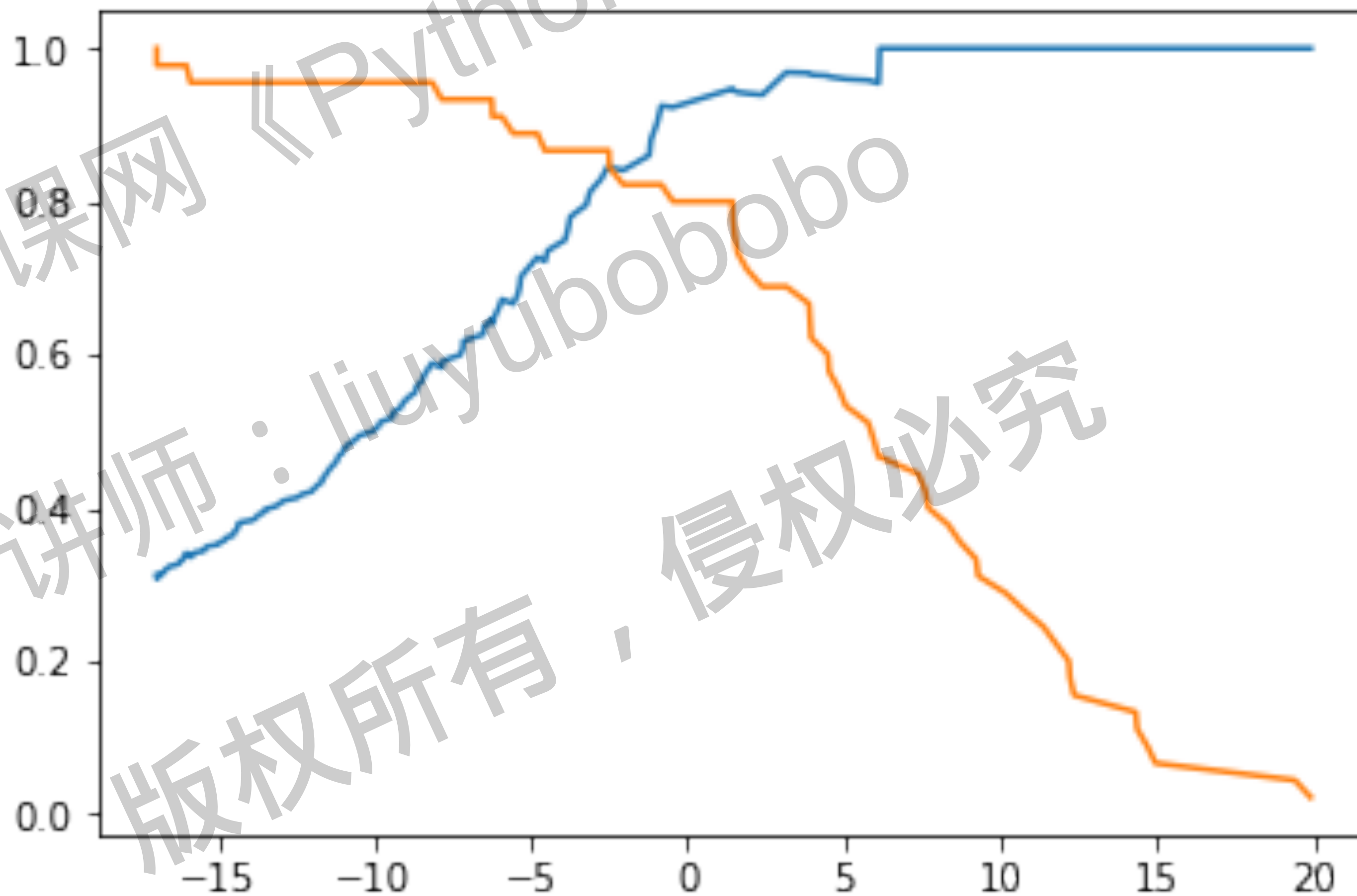
讲师：liuyubobobo

版权所有，侵权必究

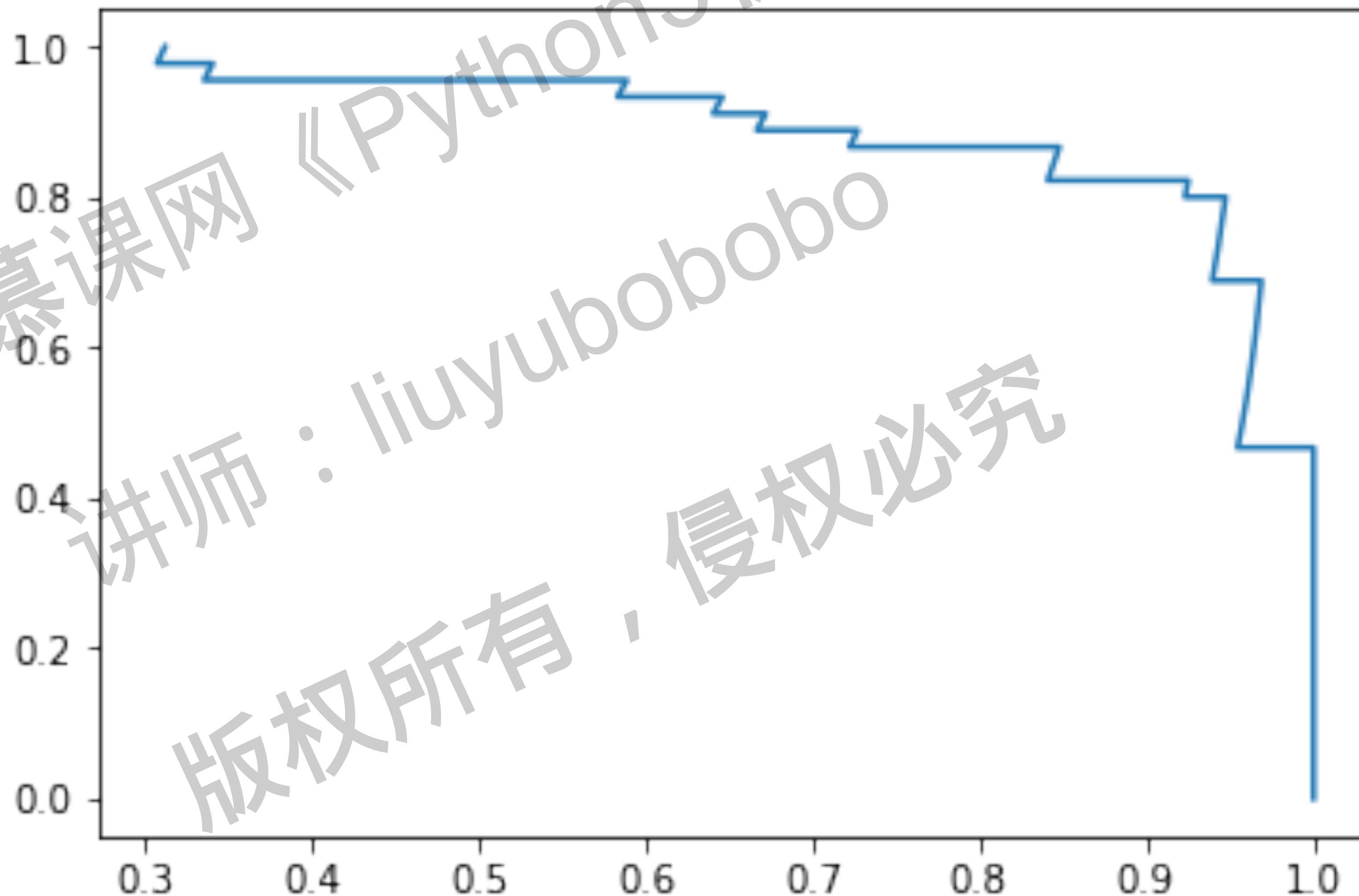
实践：Precision-Recall曲线

讲师：liuyubobobo
版权所有，侵权必究

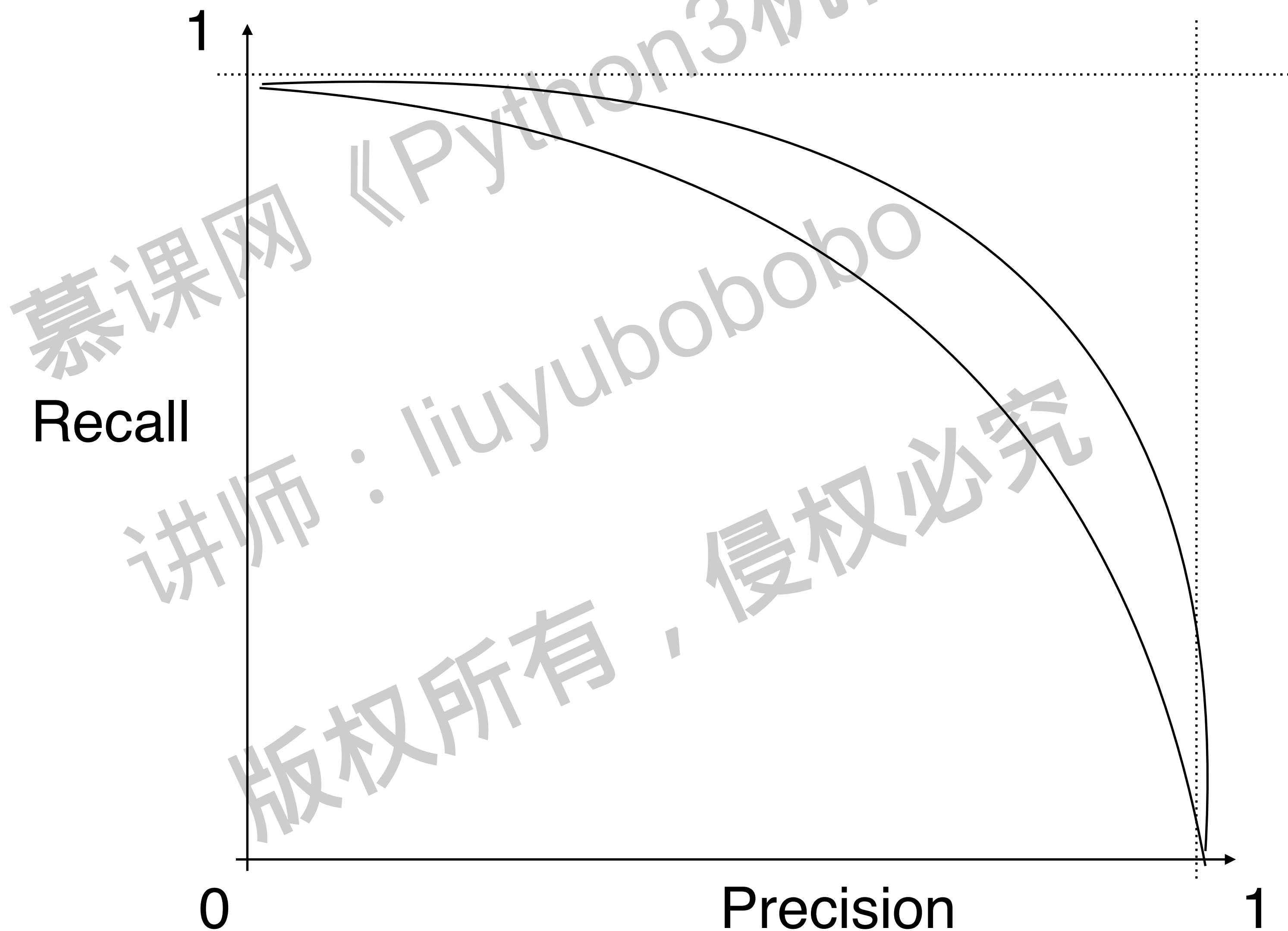
Precision-Recall 曲线



Precision-Recall 曲线



Precision-Recall 曲线



慕课网《Python3机器学习》

ROC 曲线

讲师：liuyuboboo

版权所有，侵权必究

ROC 曲线

Receiver Operation Characteristic Curve

描述TPR和FPR之间的关系

TPR

真实 \ 预测	0	1
0	9978 TN	12 FP
1	2 FN	8 TP

召回率

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

TPR = Recall

FPR

真实 \ 预测	0	1
0	9978 TN	12 FP
1	2 FN	8 TP

$$FPR = \frac{FP}{TN + FP}$$

TPR和FPR

真实 \ 预测	0	1
0	9978 TN	12 FP
1	2 FN	8 TP

$$FPR = \frac{FP}{TN + FP}$$

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

TPR和FPR的关系

FPR: $2/6=0.33$

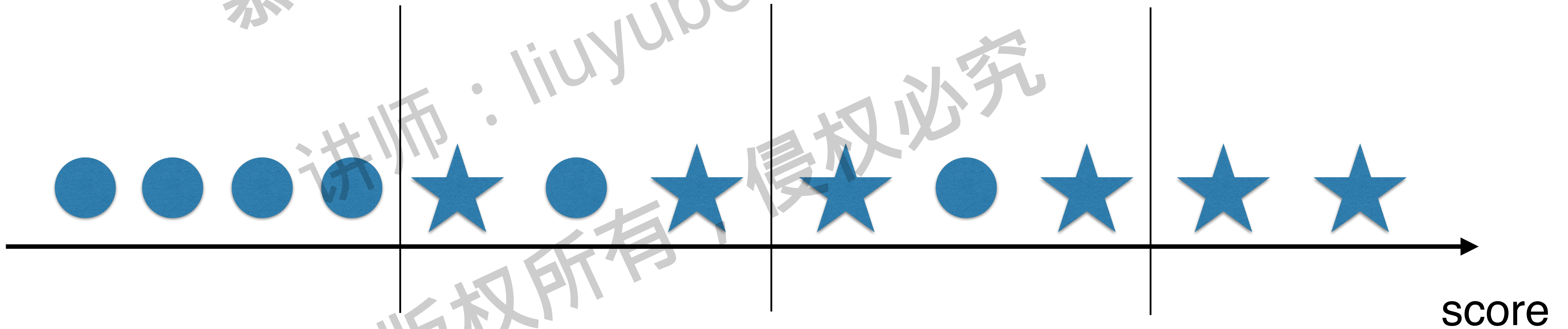
FPR: $1/6=0.16$

FPR: $0/6=0.00$

TPR: $6/6=1.00$

TPR: $4/6=0.67$

TPR: $2/6=0.33$



慕课网《Python3机器学习》

实践：实现TPR, FPR

讲师：liuyubobobo

版权所有，侵权必究

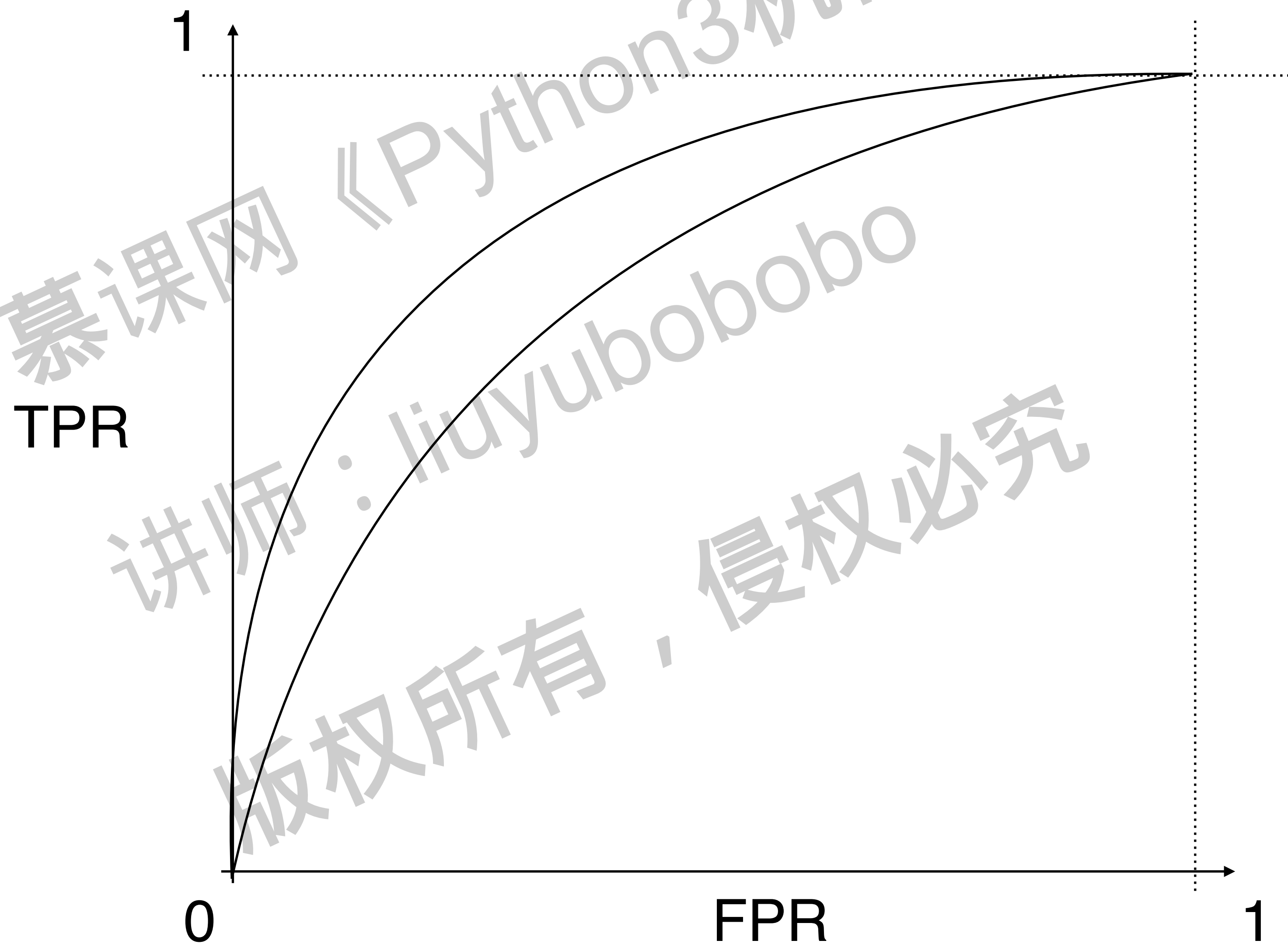
慕课网《Python3机器学习》

实践：实现ROC 曲线

讲师：liuyubobobo

版权所有，侵权必究

ROC曲线



慕课网《Python3机器学习》

多分类问题中的混淆矩阵

讲师：liuyubobobo

版权所有，侵权必究

实践：多分类问题中的混淆矩阵

讲师：liuyuboboo
版权所有，侵权必究

其他

欢迎大家关注我的个人公众号：是不是很酷



Python 3 玩儿转机器学习

讲师：liuyubobobo

版权所有 侵权必究
liuyubobobo