

法律声明

- 本课件包括：演示文稿，示例，代码，题库，视频和声音等，小象学院拥有完全知识产权的权利；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或机构不得盗版、复制、仿造其中的创意，我们将保留一切通过法律手段追究违反者的权利。



关注 **小象学院**



模型调参及评估

--Robin

模型评价指标

评估模型

- 不同的应用有着不同的目标，不同的评价指标
- **准确率（accuracy）**是最常见的一种
- 但是准确率越高，模型就越好么？
- 假设，在1000个样本中，有999个正样本，1个负样本（不均衡数据集）
如果全部预测正样本，就可以得到**准确率99.9%**！这样的场景有：信用卡欺诈检测，离职员工检测等。
- 有些任务更关心的是某个类的准确率，而非整体的准确率

模型评价指标

- 真正例(TP), 预测值是1, 真实值是1。被正确分类的正例样本。
- 假正例(FP), 预测值是1, 但真实值是0
- 真反例(TN), 预测值是0, 真实值是0
- 假反例(FN), 预测值是0, 但真实值是1。
- **TPR(Recall, 召回率)**: $TP/(TP + FN)$, 表示检测率
- **Precision(精确率)**: $TP/(TP + FP)$
- **FPR**: $FP/(TN+FP)$, 在所有实际值是0的样本中, 被错误地预测为1的比例。

		Prediction	
		Positive	Negative
Ground truth	Positive	True positive (TP)	False negative (FN)
	Negative	False positive (FP)	True negative (TN)

模型评价指标

- **F1值**

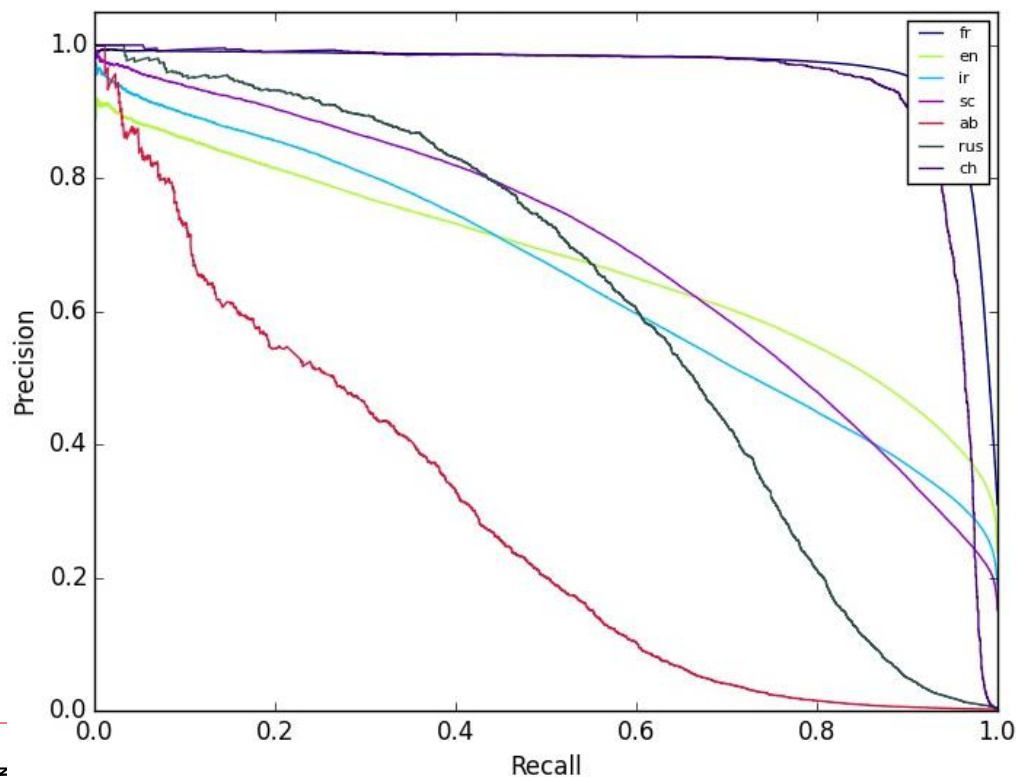
- 将召回率和精确率用一个数值表示

$$F_1 = 2 \cdot \frac{1}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}.$$

- sklearn.metrics中包含常用的评价指标
 - accuracy_score
 - precision_score
 - recall_score
 - f1_score

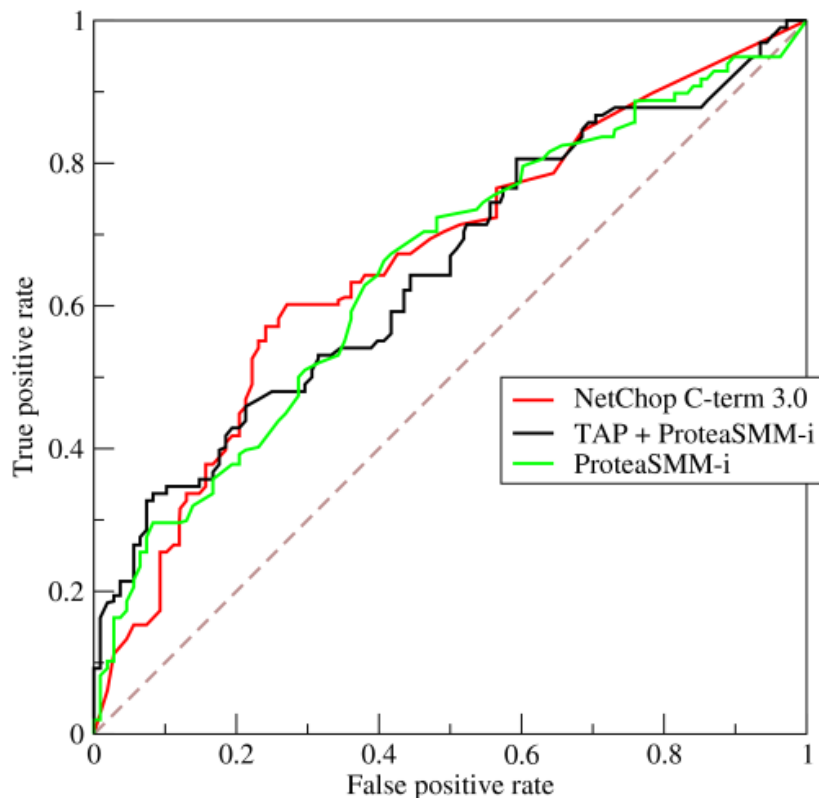
模型评价指标

- **Precision-Recall Curve (PR曲线)**
- x轴: recall, y轴: precision (可交换)
- 右上角是“最理想”的点, precision=1.0, recall=1.0
- `sklearn.metrics.precision_recall_curve()`



模型评价指标

- Receiver Operating Characteristic Curve (ROC曲线)
- x轴: FPR, y轴: TPR
- 左上角是“最理想”的点, $FPR=0.0$, $TPR=1.0$
- `sklearn.metrics.roc_curve()`

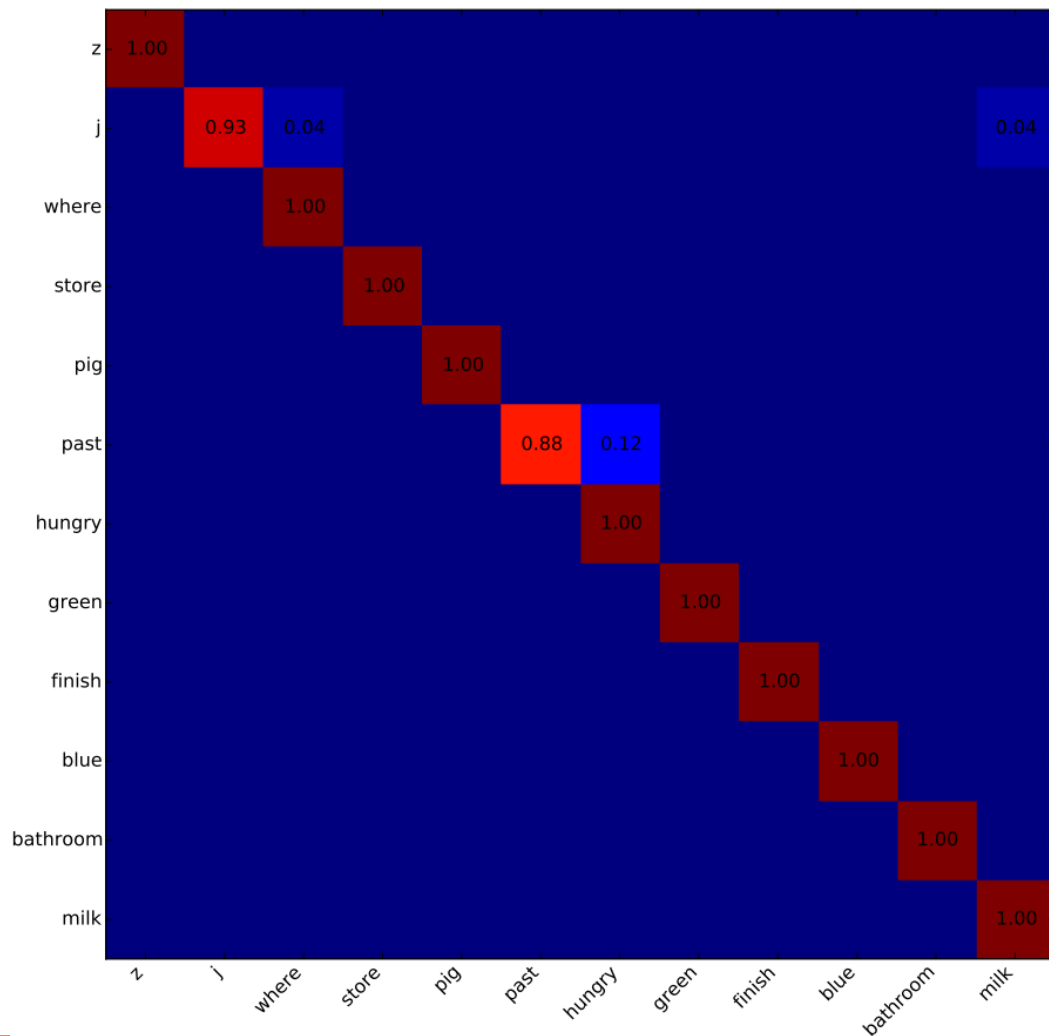


AUC的值就是ROC曲线下的面积

- AUC在0~1之间
- $0.5 < AUC < 1$, 优于随机猜测。这个分类器（模型）妥善设定阈值的话，能有预测价值。
- $AUC = 0.5$, 跟随机猜测一样（例：丢铜板），模型没有预测价值。
- $AUC < 0.5$, 比随机猜测还差；但只要总是反预测而行，就优于随机猜测。

模型评价指标

- 混淆矩阵 (**confusion matrix**)
- 可用于多分类模型的评价
- `sklearn.metrics.confusion_matrix()`



模型评价指标

- 回归模型中常用的评价指标
- `sklearn.metrics.r2_score()`
- `sklearn.metrics.mean_absolute_error()`
- `sklearn.metrics.mean_squared_error()`
- `sklearn.metrics.median_absolute_error()`

- 更多模型评价指标请参考：

http://scikit-learn.org/stable/modules/model_evaluation.html

联系我们

小象学院：互联网新技术在线教育领航者

— 微信公众号：**小象学院**

