

# 机器学习与数据挖掘-HW2

## —Evaluation Metrics—

19335253 葉珺明

### 1 MAP@K, MRR@K

#### 1.a Compute AP@5, AP@10, RR@5, RR@10

AP@K: Average of P@K; RR@K: Reciprocal Rank =  $\frac{1}{N}$  (N, rank position)

(高亮表示relevant)

- Ranking1:

R1	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
P	1.00	0.50	0.67	0.75	0.60	0.67	0.71	0.63	0.56	0.60

$$AP@5 : (1 + \frac{2}{3} + \frac{3}{4})/3 \approx 0.8056$$

$$AP@10 : (1 + \frac{2}{3} + \frac{3}{4} + \frac{4}{6} + \frac{5}{7} + \frac{6}{10})/6 \approx 0.7329$$

$$RR@5 : 1$$

$$RR@10 : 1$$

- Ranking2:

R2	d3	d8	d7	d1	d2	d4	d5	d9	d10	d6
P	0	0.50	0.33	0.25	0.20	0.17	0.14	0.25	0.22	0.20

$$AP@5 : \frac{1}{2} = 0.5$$

$$AP@10 : (\frac{1}{2} + \frac{2}{8})/2 = 0.375$$

$$RR@5 : \frac{1}{2} = 0.5$$

$$RR@10 : \frac{1}{2} = 0.5$$

- Ranking3:

R3	d7	d6	d5	d3	d2	d1	d9	d10	d4	d8
P	0	0	0.33	0.25	0.20	0.17	0.29	0.25	0.22	0.30

$$AP@5 : \frac{1}{3} \approx 0.3333$$

$$AP@10 : (\frac{1}{3} + \frac{2}{7} + \frac{3}{10})/3 \approx 0.3063$$

$$RR@5 : \frac{1}{3} \approx 0.3333$$

$$RR@10 : \frac{1}{3} \approx 0.3333$$

## 1.b Compute MAP@5, MAP@10, MRR@5, MRR@10

MAP@K: Mean AP@K; MRR@K: Mean RR@K

$$MAP@5 : (0.8056 + 0.5 + 0.3333)/3 = 0.5463$$

$$MAP@10 : (0.7329 + 0.375 + 0.3063)/3 = 0.4714$$

$$MRR@5 : (1 + 0.5 + 0.3333)/3 = 0.6111$$

$$MRR@10 : (1 + 0.5 + 0.3333)/3 = 0.6111$$

## 2 P@K, R@K, NDCG@K

### 2.a Compute P@5, p@10

$$P@5 : \frac{4}{5} = 0.8$$

$$P@10 : \frac{7}{10} = 0.7$$

### 2.b Compute R@5, R@10

$$R@5 : \frac{4}{7} \approx 0.5714$$

$$R@10 : \frac{7}{7} = 1$$

## 2.c Maximizes P@5

rank	1	2	3	4	5	6	7	8	9	10
docID	51	301	75	321	38	412	101	21	521	331
BN Rel	1	1	1	1	1	1	1	0	0	0

## 2.d Maximizes P@10

rank	1	2	3	4	5	6	7	8	9	10
docID	51	301	75	321	38	412	101	21	521	331
BN Rel	1	1	1	1	1	1	1	0	0	0

## 2.e Maximizes R@5

rank	1	2	3	4	5	6	7	8	9	10
docID	51	301	75	321	38	412	101	21	521	331
BN Rel	1	1	1	1	1	1	1	0	0	0

## 2.f Maximizes R@10

rank	1	2	3	4	5	6	7	8	9	10
docID	51	301	75	321	38	412	101	21	521	331
BN Rel	1	1	1	1	1	1	1	0	0	0

## 2.g Find a query-specific method

R-Precision可以满足该要求，R-Precision：计算序列中前R个位置文献的准确率，R为查询的相关文档的总数。

计算方法为：记R个检索文档中有r个是相关的，那么R-Precision为  $\frac{r}{R}$

## 2.h AP and MAP

$$AP: (1 + \frac{2}{2} + \frac{3}{4} + \frac{4}{5} + \frac{5}{6} + \frac{6}{8} + \frac{7}{10})/7 \approx 0.8333$$

AP: Average Precision. 把所有相关文档的P@K求平均，计算对象是一个查询中的相关文档

MAP: Mean average precision. 将所有的AP求平均，计算对象是多个查询的AP

## 2.i Maximizes AP

rank	1	2	3	4	5	6	7	8	9	10
docID	51	301	75	321	38	412	101	21	521	331
BN Rel.	1	1	1	1	1	1	1	0	0	0

## 2.j Compute $DCG_5$

Rank	1	2	3	4	5
$Rel_i$	4	1	0	3	4

$$DCG_5 = \sum_{i=1}^5 \frac{rel_i}{\log_2(i+1)} = 4 + \frac{1}{\log_2 3} + 0 + \frac{3}{\log_2 5} + \frac{4}{\log_2 6} \approx 7.4704$$

## 2.k Compute $NDCG_5$

### 2.k.(i) ideal top-5 ranking

$IDCG$  :

Rank	1	2	3	4	5	6	7	8	9	10
$Rel_i$	4	4	3	2	1	1	1	0	0	0
docID	51	321	75	101	301	38	412	21	521	331

### 2.k.(ii) Compute $IDCG_5$

$$IDCG_5 = \sum_{i=1}^5 \frac{rel_i}{\log_2(i+1)} = 4 + \frac{4}{\log_2 3} + \frac{3}{\log_2 4} + \frac{2}{\log_2 5} + \frac{1}{\log_2 6} \approx 9.2719$$

### 2.k.(iii) Compute $NDCG_5$

$$NDCG_5 = \frac{DCG_5}{IDCG_5} = \frac{7.4704}{9.2719} = 0.8057$$

## 2.j Other evaluation metrics

Other evaluation metrics: 平均倒数排名 (Reciprocal Rank)

Consider rank position  $k$ , in this table, Reciprocal Ranking score is:

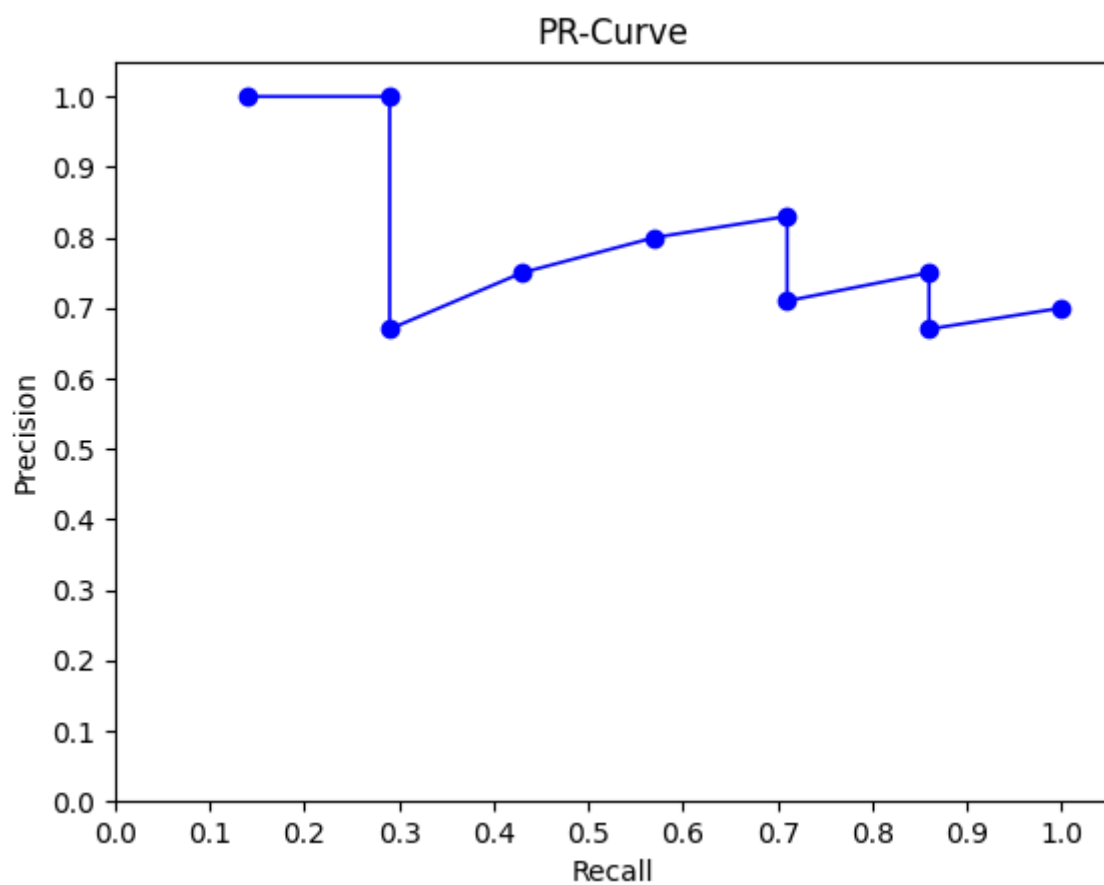
$$RR = \frac{1}{1} = 1$$

**MRR**(Mean Reciprocal Rank) can be used across multiple queries.

## 3 Precision-Recall Curves

Rank	1	2	3	4	5	6	7	8	9	10
Rel.	1	1	0	1	1	1	0	1	0	1
P	1.00	1.00	0.67	0.75	0.80	0.83	0.71	0.75	0.67	0.70
R	0.14	0.29	0.29	0.43	0.57	0.71	0.71	0.86	0.86	1.00

**PR-Curve for E2 query:**



## 4 Other Evaluation Metrics

- 按照学习类型分类：

学习类型	性能度量
分类	accuracy、precision、recall、F1 score、ROC、AUC等
回归	MAE、MSE等

### 4.1 分类的性能度量

#### 4.1.1 准确率、精确率、召回率和F1

- 混淆矩阵
  - True Positive(TP): 预测为真的正样本
  - True Negative(TN): 预测为假的负样本
  - False Positive(FP): 预测为真的负样本
  - False Negative(FN): 预测为假的正样本
- Accuracy 准确率：正确预测占所有样本的比例

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- Precision 精确度：预测为真的正样本占所有预测为真的比例

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

- Recall 召回率：预测为真的正样本占所有实际正样本的比例

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

- F1 Score：基于精确率和召回率的调和均值

$$\frac{1}{F_1} = \frac{1}{2} \left( \frac{1}{P} + \frac{1}{R} \right)$$
$$F_1 = \frac{2TP}{2TP + FP + FN}$$

加权调和平均：

$$F_\beta = \frac{1}{1 + \beta^2} \left( \frac{1}{P} + \frac{\beta^2}{R} \right)$$

其中 $\beta$ 度量了召回率对精确率的相对重要性。 $\beta = 1$ 退化为F1； $\beta > 1$ 召回率有更大影响； $\beta < 1$ 精确率有更大影响。

### 4.1.2 ROC和AUC

- ROC(Receiver Operation Characteristic Curve)

在分类问题中，模型输出为样本点属于类别的概率大小，通过设定阈值将样本归类。

- TPR：预测为真的正样本占有所有实际正样本的比例；FPR：预测为真的负样本占有所有实际负样本的比例

$$TPR = \frac{TP}{TP + FN}; \quad FPR = \frac{FN}{FN + TN}$$

- 以TPR为纵坐标，FPR为横坐标绘制ROC曲线

- AUC(Area Under the Curve of ROC)

ROC曲线下的面积，即AUC，AUC值越大的分类器，正确率越高。

- $AUC = 1$ ，完美分类器，不管设定什么阈值，都能得到完美预测。
- $1 > AUC > 0.5$ ，优于随机猜测，该分类器妥善设定阈值时，能有预测价值。
- $AUC = 0.5$ ，随机猜测，没有预测价值。
- $AUC < 0.5$ ，劣于随机猜测；但只要反预测，就优于随机猜测，因此不存在  $AUC < 0.5$  的情况

ROC和AUC绘制例子：

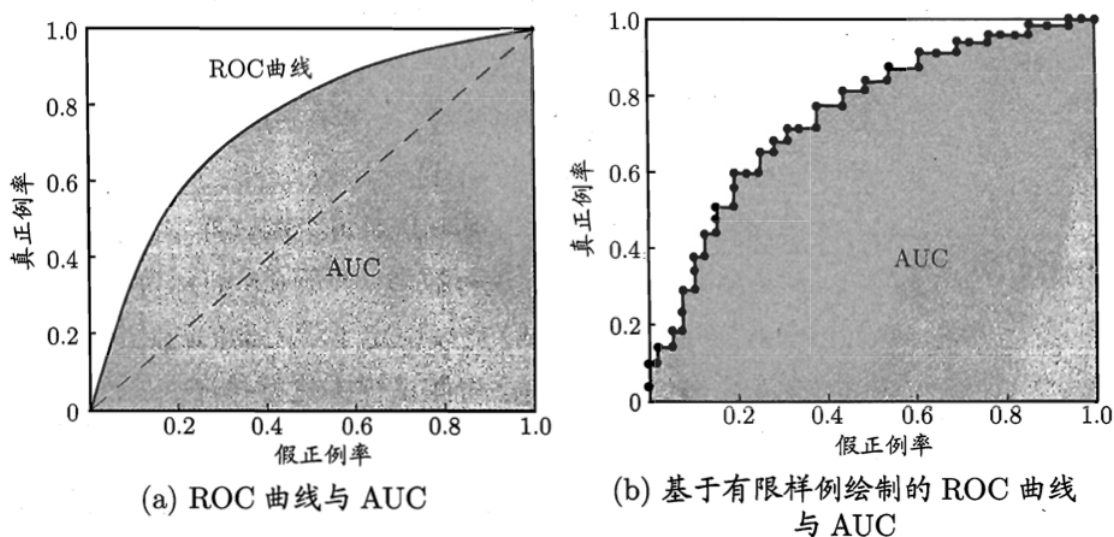


图 2.4 ROC 曲线与 AUC 示意图

### 4.1.3 代价敏感错误率与代价曲线

在评价学习模型性能时考虑不同类分类错误所造成不同损失代价的因素时，称为代价敏感错误率评估方法。

- 均等代价，分类错误的损失代价相同
- 非均等代价，分类错误的损失代价不同

以二分类任务为例，设定“代价矩阵”：

真实类别	预测类别	
	第0类	第1类
第0类	0	$cost_{01}$
第1类	$cost_{10}$	0

$cost_{ij}$ 表示错误判断时的代价，上图 $cost_{01} > cost_{10}$ ，将第0类定为正类，第1类定为反类， $D^+$ 和 $D^-$ 分别代表样例集 $D$ 的正例子集和反例子集，在非均等错误代价下，我们希望的是最小化“总体代价”，则“代价敏感”错误率为：

$$E(f; D; cost) = \frac{1}{m} \left( \sum_{\mathbf{x}_i \in D^+} \mathbb{I}(f(\mathbf{x}_i) \neq y_i) \times cost_{01} + \sum_{\mathbf{x}_i \in D^-} \mathbb{I}(f(\mathbf{x}_i) \neq y_i) \times cost_{10} \right)$$

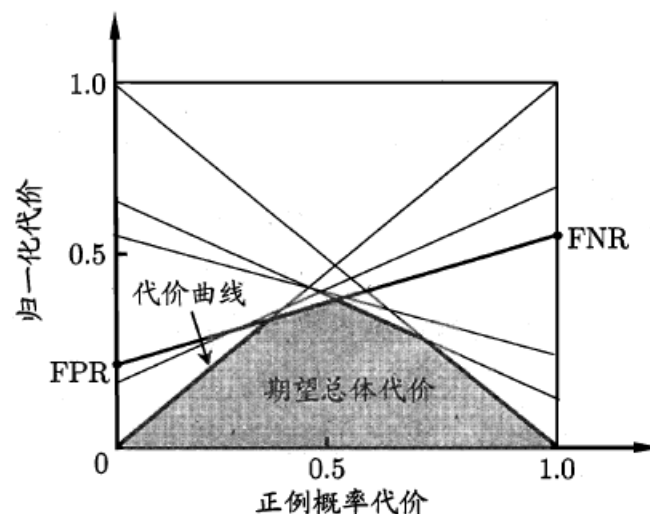
- 非均等代价下，通过“代价曲线”反映出机器学习的期望总体代价  
横轴为正例概率：其中 $p$ 表示正例的概率。

$$P(+)\text{cost} = \frac{p \times cost_{01}}{p \times cost_{01} + (1 - p) \times cost_{10}},$$

纵轴是取值为[0,1]的归一化代价： $FNR = 1 - TPR$ 是假反例率。

$$cost_{norm} = \frac{FNR \times p \times cost_{01} + FPR \times (1 - p) \times cost_{10}}{p \times cost_{01} + (1 - p) \times cost_{10}}$$

绘制代价曲线：设ROC曲线上一点的坐标为(TPR, FPR)，则可相应计算出FNR，然后在代价平面上绘制一条从(0, FPR)到(1, FNR)的线段，线段下的面积即表示了该条件下的期望总体代价；如此将ROC曲线上的每个点转化为代价平面上的一条线段，然后取所有线段的下界，围成的面积即为在所有条件下学习器的期望总体代价，如图所示：





## 4.2 回归问题的性能度量

- MAE(Mean Absolution Error) 平均绝对误差

$$MAE(m) = \frac{1}{m} \sum_{i=1}^m |f(x_i) - f'(x_i)|$$

- MSE(Mean Squared Error) 平均平方误差

$$MSE(m) = \frac{1}{m} \sum_{i=1}^m (f(x_i) - f'(x_i))^2$$