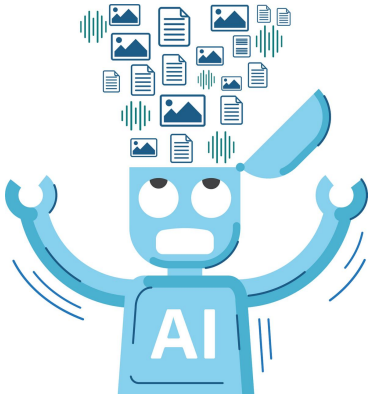
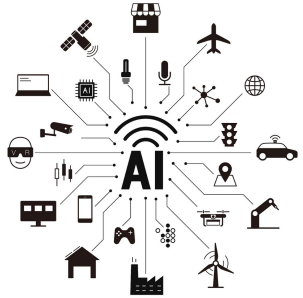
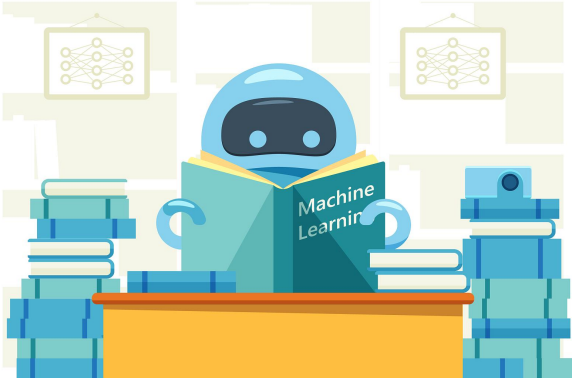


机器学习
Machine Learning



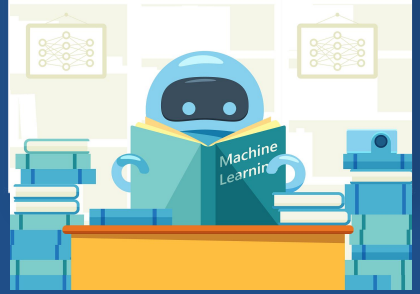
软 件 学 院　罗 昕
luoxin@sdu.edu.cn



山东大学软件学院
SCHOOL OF SOFTWARE, SHANDONG UNIVERSITY

Machine Learning

机器学习



Supplementary Materials Evaluation

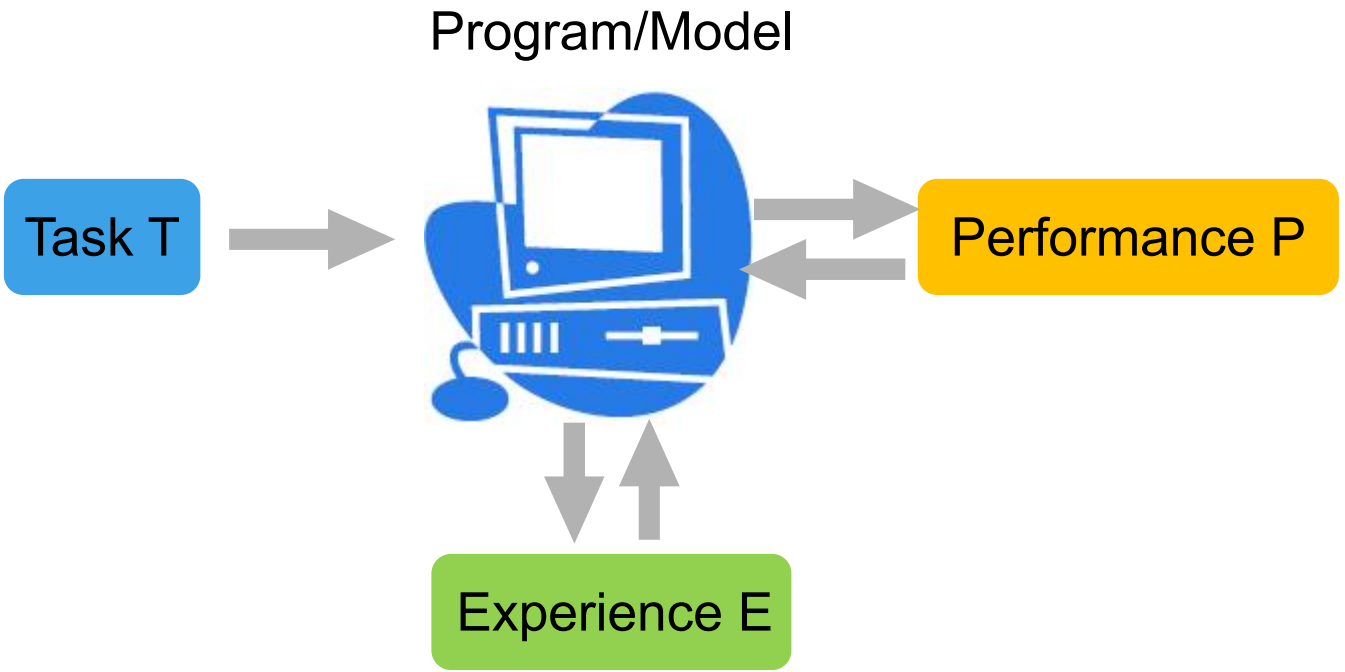
软件学院 罗昕

luoxin@sdu.edu.cn

软件学院办公楼-425

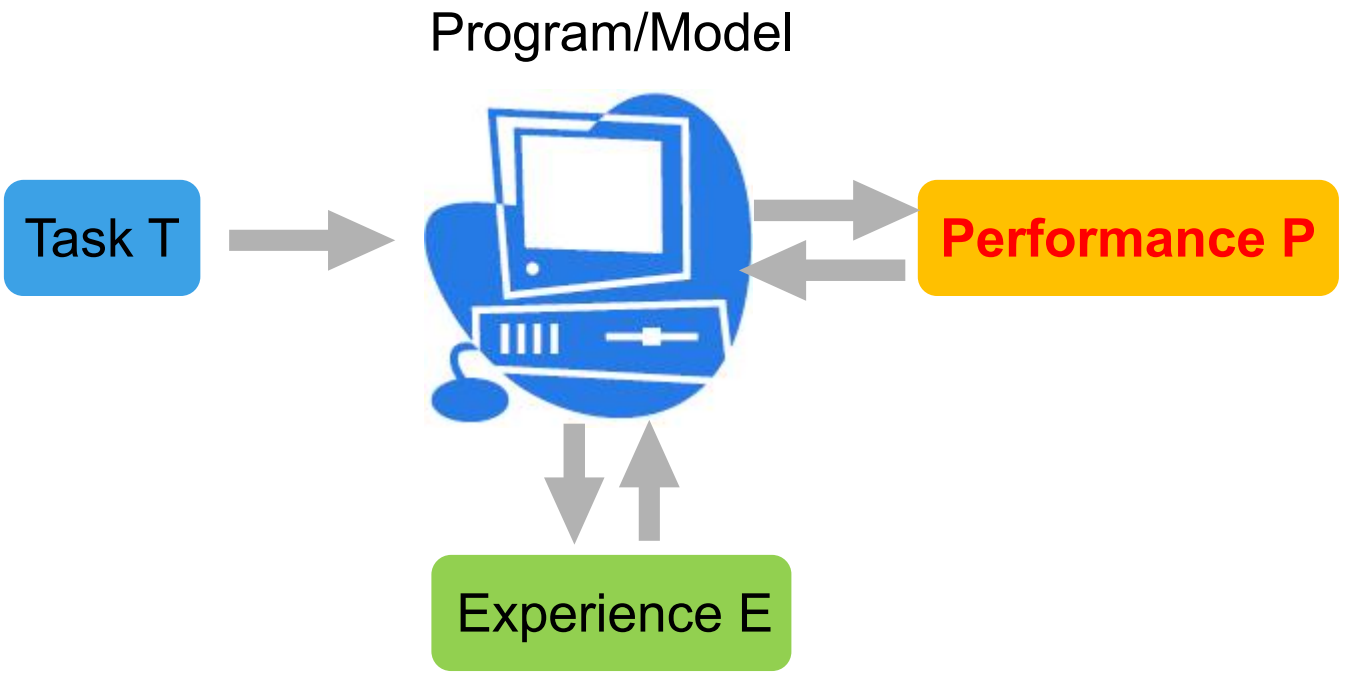
What is Machine Learning?

MIMA



What is Machine Learning?

MIMA



salmon	sea bass
鲑鱼	鲈鱼

正例与负例

MIMA

- 喜欢 salmon， 则为正例

> 正例（Positives）：你所关注的识别目标就是正例。
>
> 负例（Negatives）：正例以外的就是负例。

正例与负例

Fig. 1. Sample images from the MNIST dataset

正例与负例

TP FN TN FP

MIMA

符号	简称	含义	之和
TP（True Positives）	真正例	识别对了的正例（实际是正例）	实际的正例数量
FN（False Negatives）	伪负例	识别错了的负例（实际是正例）	
TN（True Negatives）	真负例	识别对了的负例（实际是负例）	实际的负例数量
FP（False Positives）	伪正例	识别错了的正例（实际是负例）	

单选题 1分

设置

MIMA

识别出的正例

A TP+FP

B TP+FN

符号	简称	含义	之和
TP（True Positives）	真正例	识别对了的正例（实际是正例）	实际的正例数量
FN（False Negatives）	伪负例	识别错了的负例（实际是正例）	
TN（True Negatives）	真负例	识别对了的负例（实际是负例）	实际的负例数量
FP（False Positives）	伪正例	识别错了的正例（实际是负例）	

提交

设置

MIMA

识别出的负例

A TN+FN

B TN+FP

符号	简称	含义	之和
TP（True Positives）	真正例	识别对了的正例（实际是正例）	实际的正例数量
FN（False Negatives）	伪负例	识别错了的负例（实际是正例）	
TN（True Negatives）	真负例	识别对了的负例（实际是负例）	实际的负例数量
FP（False Positives）	伪正例	识别错了的正例（实际是负例）	

提交

单选题 1分

设置

MIMA

总共识别的样本数是TP+FN+TN+FP吗?

A 是

B 不是

符号	简称	含义	之和
TP（True Positives）	真正例	识别对了的正例（实际是正例）	实际的正例数量
FN（False Negatives）	伪负例	识别错了的负例（实际是正例）	
TN（True Negatives）	真负例	识别对了的负例（实际是负例）	实际的负例数量
FP（False Positives）	伪正例	识别错了的正例（实际是负例）	

提交

设置

MIMA

识别对的样本

A TP+TN

B FP+FN

符号	简称	含义	之和
TP（True Positives）	真正例	识别对了的正例（实际是正例）	实际的正例数量
FN（False Negatives）	伪负例	识别错了的负例（实际是正例）	
TN（True Negatives）	真负例	识别对了的负例（实际是负例）	实际的负例数量
FP（False Positives）	伪正例	识别错了的正例（实际是负例）	

提交

单选题 1分

设置

MIMA

识别错的样本

A FP+FN

B TP+TN

符号	简称	含义	之和
TP（True Positives）	真正例	识别对了的正例（实际是正例）	实际的正例数量
FN（False Negatives）	伪负例	识别错了的负例（实际是正例）	
TN（True Negatives）	真负例	识别对了的负例（实际是负例）	实际的负例数量
FP（False Positives）	伪正例	识别错了的正例（实际是负例）	

提交

机器学习中的评价指标

MIMA

- 1 正确率（Accuracy）
- 2 错误率（Error-rate）
- 3 精度（Precision）
- 4 召回率（Recall）
- 5 精度-召回率曲线（PR曲线）
- 6 AP（Average Precision）值
- 7 mAP（Mean Average Precision）值
- 8 综合评价指标F-Measure
- 9 ROC曲线与AUC
- 10 IoU（Intersection-over-Union）指标
- 11 Top1与TopK

1 正确率（Accuracy）

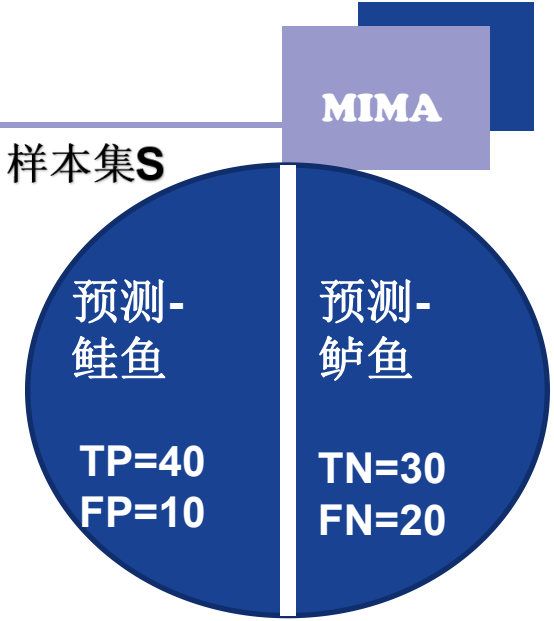
- 正确率（Accuracy）：也即准确率，识别对了的正例（TP）与负例（TN）占总识别样本的比例。

- **A=(TP+ TN)/S**

在例子中，TP+ TN =70，S= 100，正确率：

A=70/100=0.7

- 通常来说，正确率越高，模型性能越好。



TP 真的鲑鱼
（鲑鱼正确识别为鲑鱼）
FP 假的鲑鱼
（鲈鱼错误识别为鲑鱼）
TN 真的鲈鱼
（鲈鱼正确识别为鲈鱼）
FN 假的鲈鱼
（鲑鱼错误识别为鲈鱼）

2 错误率（Error-rate）

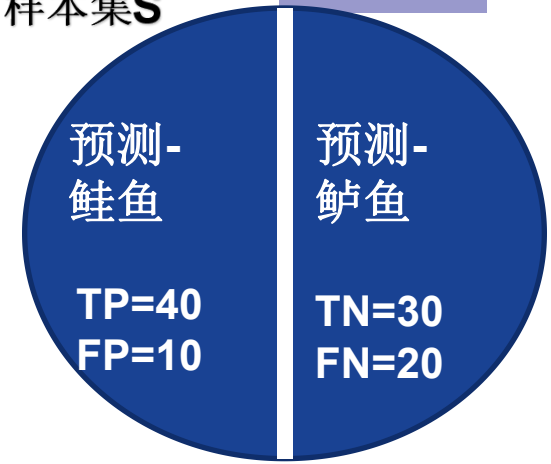
MIMA

- 错误率（Error-rate）：识别错了的正例（FP）与负例（FN）占总识别样本的比例。

- **E=(FP+FN)/S**

  E=30/100=0.3

- 可见，**正确率与错误率是分别从正反两方面进行评价的指标，两者数值相加刚好等于1。**正确率高，错误率就低；正确率低，错误率就高。



TP 真的鲑鱼
（鲑鱼正确识别为鲑鱼）
FP 假的鲑鱼
（鲈鱼错误识别为鲑鱼）
TN 真的鲈鱼
（鲈鱼正确识别为鲈鱼）
FN 假的鲈鱼
（鲑鱼错误识别为鲈鱼）

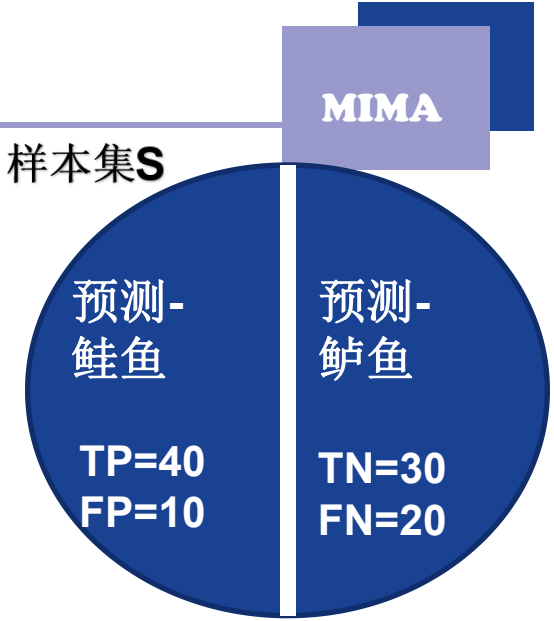
3 精度（Precision）

- 精度（Precision）：识别对了的正例（TP）占识别出的正例的比例。其中，识别出的正例等于识别对了的正例加上识别错了的正例。

- **P=TP/(TP+ FP)**

  P=40/50=0.8

- 因此，精度即为识别目标正确的比例。精度也即查准率，好比例子来说，模型查出了50个目标，但这50个目标中准确的比率有多少。

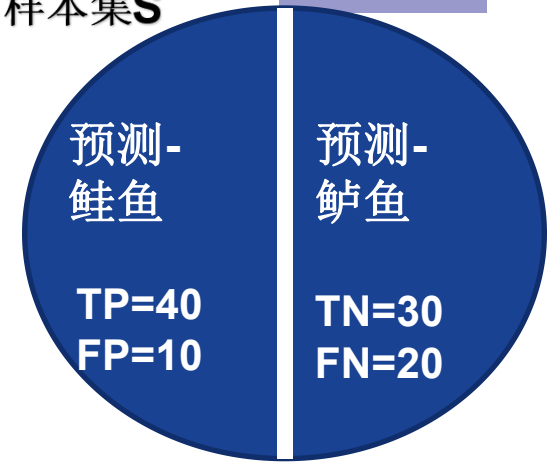


TP 真的鲑鱼
（鲑鱼正确识别为鲑鱼）
FP 假的鲑鱼
（鲈鱼错误识别为鲑鱼）
TN 真的鲈鱼
（鲈鱼正确识别为鲈鱼）
FN 假的鲈鱼
（鲑鱼错误识别为鲈鱼）

4 召回率（Recall）

- 召回率（Recall）：识别对了的正例（TP）占实际总正例的比例。其中，实际总正例等于识别对了的正例加上识别错了的负例（真正例+伪负例）。
- **R=TP/(TP+ FN)**

  R=40/60=0.67



TP 真的鲑鱼
（鲑鱼正确识别为鲑鱼）
FP 假的鲑鱼
（鲈鱼错误识别为鲑鱼）
TN 真的鲈鱼
（鲈鱼正确识别为鲈鱼）
FN 假的鲈鱼
（鲑鱼错误识别为鲈鱼）

- 在一定意义上来说，召回率也可以说是“找回率”，也就是在实际的60个目标中，找回了40个，找回的比例即为：40/60。同时，召回率也即查全率，即在实际的60个目标中，有没有查找完全，查找到的比率是多少。

Precision & Recall

MIMA

- **P=TP/(TP+ FP)** Precision
- **R=TP/(TP+ FN)** Recall
- 从公式可以看出，精度与召回率都与TP值紧密相关，TP值越大，精度、召回率就越高。理想情况下，我们希望精度、召回率越高越好。但单独的高精度或高召回率，都不足以体现模型的高性能。

高精度 却 低性能模型

MIMA

- 精度P为100%
- 但是识别给出的200个负例全部都错误（都是伪负例），错误率非常高，这样的模型性能其实非常低。

类别	数量	真假情况	符号	精度与错误率
正例	50	50	TP	P=TP/(TP+ FP)=50/50=100%
		0	FP	E=(FP+FN)/S=200/250=80%
负例	200	0	TN	
		200	FN	

高召回 却 低性能模型

MIMA

- 召回R为100%
- 但同时，计算得出模型识别结果的错误率E也很高，高达91%，所以这个模型性能也很低，基本不可靠。

类别	数量	真假情况	符号	召回率与错误率
正例	110	10	TP	R=TP/(TP+ FN)=10/10=100%
		100	FP	E=(FP+FN)/S=100/110=91%
负例	0	0	TN	
		0	FN	

5 精度-召回率曲线（PR曲线）

MIMA

- 实际中，精度与召回率是**相互影响**的。
- 通常，精度高时，召回率就往往偏低，而召回率高时，精度则会偏低。

- 这其实也很好理解，前面我们说了，精度即**查准率**，召回率即**查全率**，要想查得精准（一查一个准），即模型给出的目标都正确，那就得提高阈值门槛，阈值一提高，符合要求的目标就会减少，那必然会导致漏网之鱼增多，召回率降低。

5 精度-召回率曲线（PR曲线）

MIMA

- 实际中，精度与召回率是**相互影响**的。
- 通常，精度高时，召回率就往往偏低，而召回率高时，精度则会偏低。

- 相反，若想召回率高，没有漏网之鱼（目标都找到），就要降低阈值门槛，才能把所有目标收入囊中，与此同时会揽入一些伪目标，从而导致精度降低

5 PR曲线

MIMA

序号	置信度分数 (Score)	阈值（T=0.6）	阈值（T=0.5）	真实属性
1	0.86	1	1	1
2	0.97	1	1	1
3	0.99	1	1	1
4	0.85	1	1	1
5	0.78	1	1	1
6	0.72	1	1	0
7	0.74	1	1	0
8	0.63	1	1	1
9	0.58	0	1	1
10	0.55	0	1	0
11	0.48	0	0	0
12	0.46	0	0	0
13	0.32	0	0	0
14	0.22	0	0	0
15	0.19	0	0	0

阈值	TP	FP	FN
T=0.6	6	2	1
T=0.5	7	3	0

R	P
TP/(TP+ FN)=0.86	TP/(TP+ FP)=0.75
TP/(TP+ FN)=1	TP/(TP+ FP)=0.7

5 PR曲线

MIMA

序号	置信度分数 (Score)	阈值（T=0.6）	阈值（T=0.5）	真实属性
1	0.86	1		1
2	0.97	1		1
3	0.99	1		1
4	0.85	1		1
5	0.78	1		1
6	0.72	1		0
7	0.74	1		0
8	0.63	1		1
9	0.58	0		1
10	0.55	0		0
11	0.48	0		0
12	0.46	0		0
13	0.32	0		0
14	0.22	0		0
15	0.19	0		0

阈值	TP	FP	FN
T=0.6	6	2	1

R	P
TP/(TP+ FN)=0.86	TP/(TP+ FP)=0.75

阈值 为 0.6

（R=0.86，P=0.75）

5 PR曲线

MIMA

序号	置信度分数 (Score)	阈值（T=0.6）	阈值（T=0.5）	真实属性
1	0.86		1	1
2	0.97		1	1
3	0.99		1	1
4	0.85		1	1
5	0.78		1	1
6	0.72		1	0
7	0.74		1	0
8	0.63		1	1
9	0.58		1	1
10	0.55		1	0
11	0.48		0	0
12	0.46		0	0
13	0.32		0	0
14	0.22		0	0
15	0.19		0	0

阈值	TP	FP	FN
T=0.5	7	3	0

R	P
TP/(TP+ FN)=1	TP/(TP+ FP)=0.7

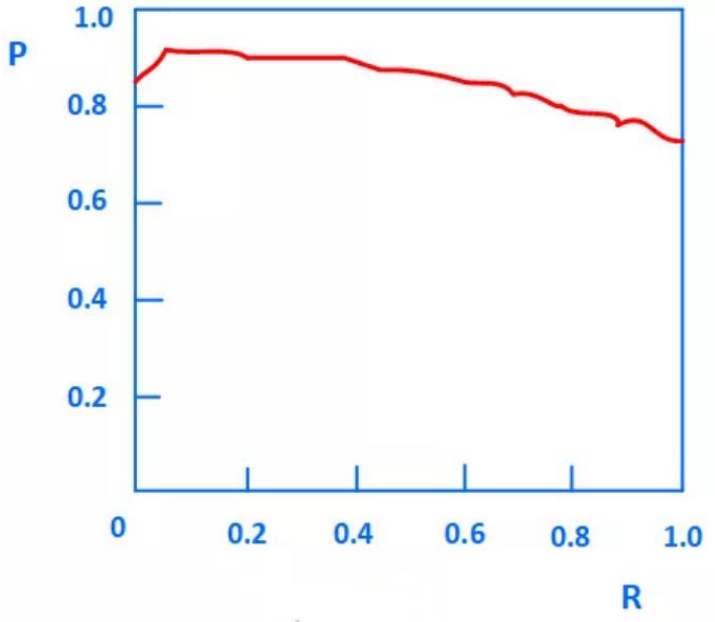
阈值 为 0.5

（R=1，P=0.7）

5 精度-召回率曲线（PR曲线）

MIMA

- 设定的阈值不同，得出的召回率（R）和精度（P）也不相同。
- 如果取多个不同的阈值，就可以得到多组（R，P）。



6 AP（Average Precision）值

MIMA

- PR曲线下的面积称为AP（Average Precision），表示召回率从0-1的平均精度值。

- 如何计算AP呢？很显然，根据数学知识，可用积分进行计算，公式如下：

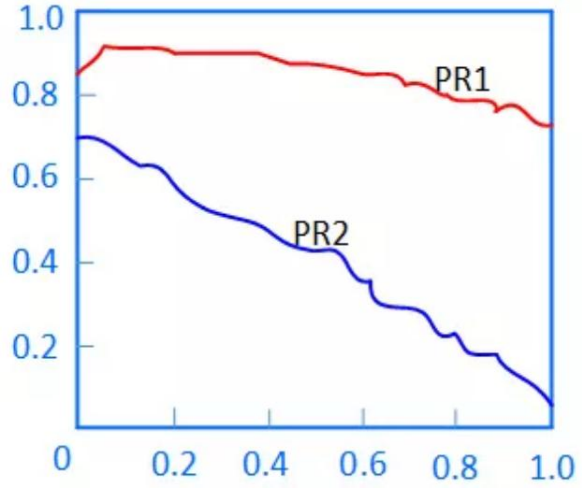
$$\mathrm{AP}=\int_0^1 p(r)dr$$

6 AP（Average Precision）值

MIMA

- AP不会大于1。

- PR曲线下的面积越大，模型性能则越好。性能优的模型应是在召回率（R）增长的同时保持精度（P）值都在一个较高的水平。



6 AP（Average Precision）值

MIMA

- 除了使用积分方法计算AP值，实际应用中，还常使用插值方法进行计算。常见的一种插值方法是：选取11个精度点值，然后计算出这11个点的平均值即为AP值。

- 怎样选取11个精度点值呢？通常先设定一组阈值,例如[0,0.1,0.2…,1], 对于R大于每一个阈值（R>0, R>0.1,…, R>1），会得到一个对应的最大精度值Pmax,这样就会得到11个最大精度值（Pmax1, Pmax2,…, Pmax11）。

- 则：AP=（Pmax1+ Pmax2+…+ Pmax11）/11

7 mAP（Mean Average Precision）值

MIMA

- AP是衡量模型在单个类别上平均精度的好坏，mAP则是衡量模型在所有类别上平均精度的好坏，每一个类别对应有一个AP，假设有n个类别，则有n个AP，分别为：AP1，AP2，…，APn, mAP就是取所有类别 AP 的平均值，即：

- mAP= (AP1+ AP2+…+ APn)/n

8 综合评价指标F-Measure

MIMA

- F-Measure又称F-Score，是召回率R和精度P的加权调和平均，顾名思义即是为了调和召回率R和精度P之间增减反向的矛盾，该综合评价指标F引入了系数α对R和P进行加权调和，表达式如下：

$$F= (\alpha^2+1)\ P.R/\alpha^2 (P+R)$$

- 而我们最常用的F1指标，就是上式中系数α取值为1的情形，即：

$$F1=2P.R/(P+R)$$

- F1的最大值为1，最小值为0。

9 ROC曲线与AUC

MIMA

- ROC(Receiver Operating Characteristic)曲线与AUC(Area Under the Curver)
- ROC曲线，也称受试者工作特征。ROC曲线与真正率（TPR，True Positive Rate）和假正率(FPR, False Positive Rate)密切相关。

- 真正率(TPR): 识别对了的正例（TP）占实际总正例的比例，实际计算值跟召回率相同。即：

$$TPR = TP/(TP + FN)$$

9 ROC曲线与AUC

MIMA

- ROC(Receiver Operating Characteristic)曲线与AUC(Area Under the Curver)
- ROC曲线，也称受试者工作特征。ROC曲线与真正率（TPR，True Positive Rate）和假正率(FPR, False Positive Rate)密切相关。

- 假正率(FPR): 识别错了的正例（FP）占实际总负例的比例。也可以说，误判的负例（实际是负例，没有判对）占实际总负例的比例。计算式如下

$$FPR = FP/(FP + TN)$$

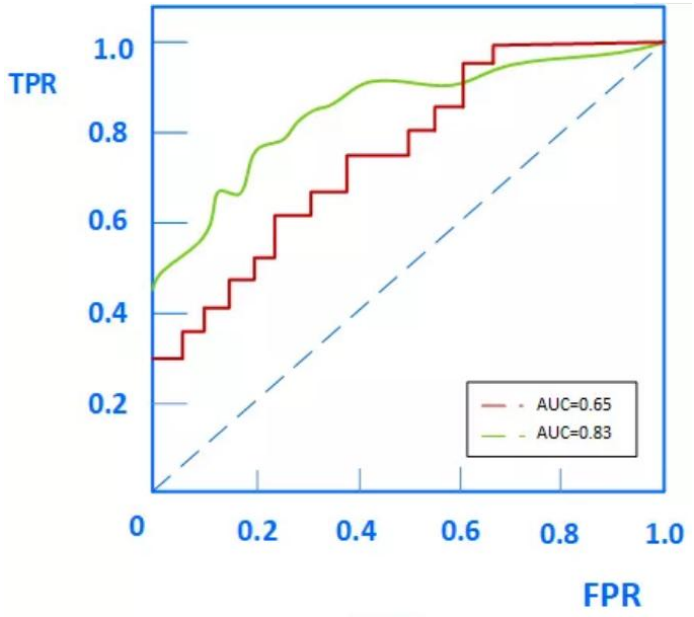
9 ROC曲线与AUC

MIMA

- 以**假正率**FPR为横轴，**真正率**TPR为纵轴，绘制得到的曲线就是ROC曲线，绘制方法与PR曲线类似。绘制得到的ROC曲线示例如下：

TPR =TP/(TP+ FN) 真正率

FPR =FP/(FP+ TN) 假正率



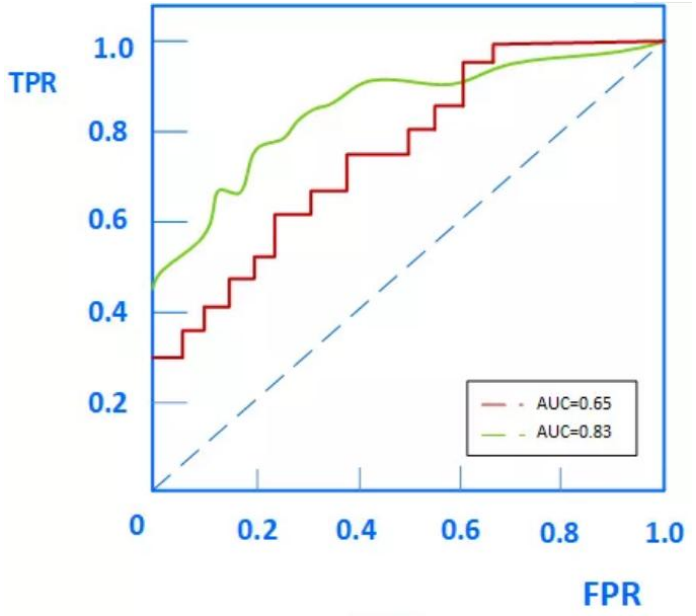
9 ROC曲线与AUC

MIMA

- ROC曲线下的面积即为AUC。面积越大性能越好。
- 绿线AUC=0.83 > 红线AUC=0.65。并且，绿线较红线更光滑。通常来说，ROC曲线越光滑，过拟合程度越小。绿线模型的整体性能要优于红线模型。

TPR =TP/(TP+ FN) 真正率

FPR =FP/(FP+ TN) 假正率



10 IoU（Intersection-over-Union）

MIMA

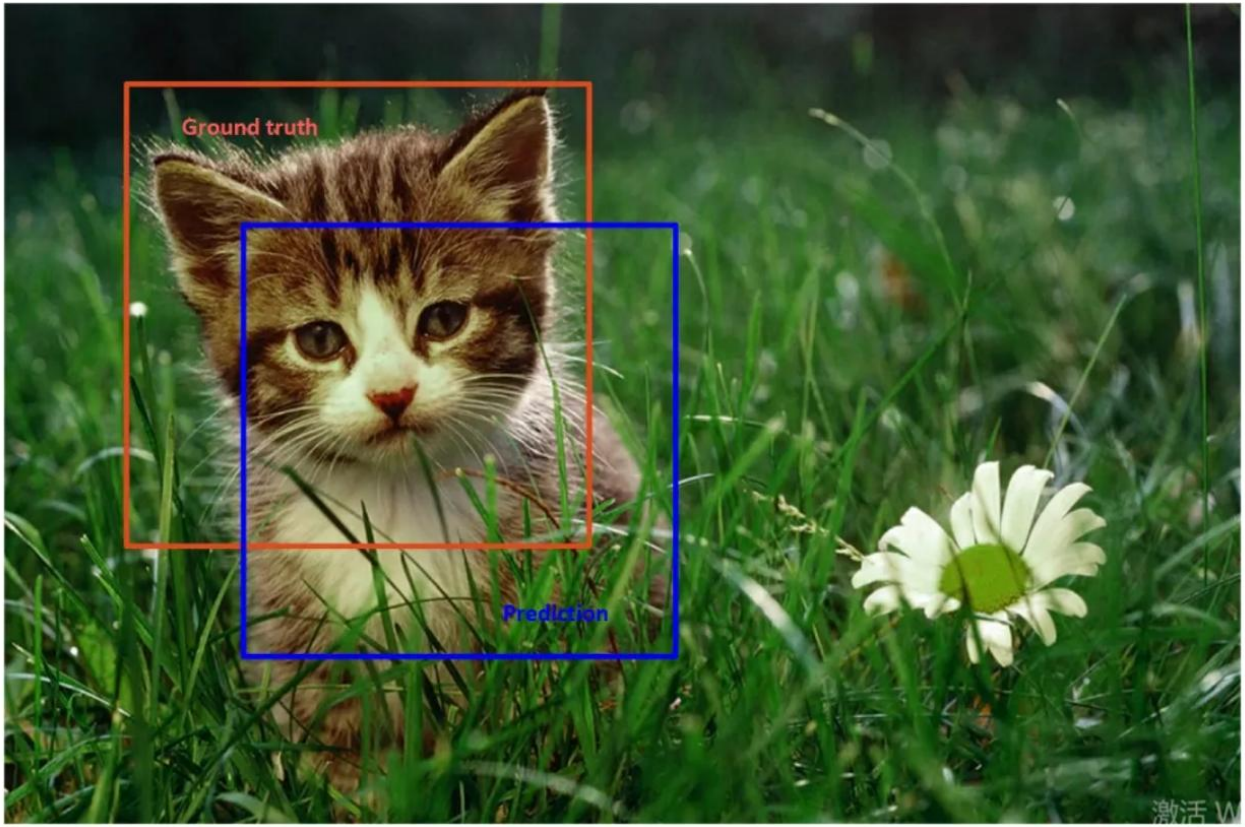
- IoU简称交并比，顾名思义数学中交集与并集的比例。假设有两个集合A与B, IoU即等于A与B的交集除以A与B的并集，表达式如下：

$$IoU = A \cap B / A \cup B$$

- 在目标检测中，IoU为预测框(Prediction)和真实框(Ground truth)的交并比。如下图所示，在关于小猫的目标检测中，**蓝色**边框为预测框(Prediction)，**红线**边框为真实框(Ground truth)。

10 IoU（Intersection-over-Union）

MIMA



蓝色边框为预测框(Prediction)
红线边框为真实框(Ground truth)

11 Top1与TopK

MIMA

- Top1：对一张图片，模型给出的识别概率中（即置信度分数），分数最高的为正确目标，则认为正确。这里的目标也就是我们说的正例。

- TopK: 对一张图片，模型给出的识别概率中（即置信度分数），分数排名前K位中包含有正确目标（正确的正例），则认为正确。

- K的取值一般可在100以内的量级，当然越小越实用。比如较常见的，K取值为5，则表示为Top5，代表置信度分数排名前5当中有一个是正确目标即可；如果K取值100，则表示为Top100，代表置信度分数排名前100当中有一个是正确目标（正确的正例）即可。可见，随着K增大，难度下降。

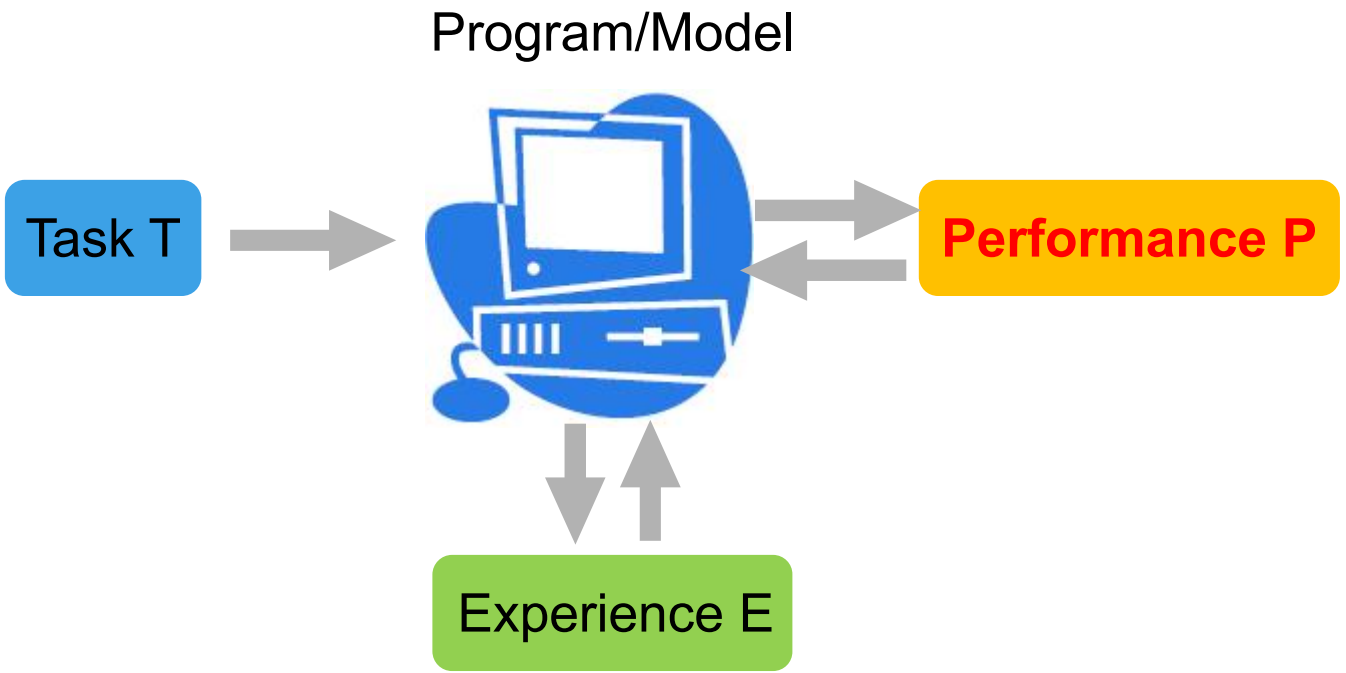
11 Top1与TopK

MIMA

- 取阈值T=0.45，排名前5的置信度分数均大于阈值，因此都识别为正例。对于Top1来说，即ID号为4的图片，实际属性却是负例，因此目标识别错误。而对于Top5来说，排名前5的置信度分数中，有识别正确的目标，即ID号为2、20的图片，因此认为正确。

ID	置信度分数(Score)	阈值（T=0.45)	真实属性
4	0.93	1	0
2	0.80	1	1
15	0.77	1	0
9	0.65	1	0
20	0.46	1	1

What is Machine Learning?



机器学习中的评价指标

MIMA

- 1 正确率（Accuracy）
- 2 错误率（Error-rate）
- 3 精度（Precision）
- 4 召回率（Recall）
- 5 精度-召回率曲线（PR曲线）
- 6 AP（Average Precision）值
- 7 mAP（Mean Average Precision）值
- 8 综合评价指标F-Measure
- 9 ROC曲线与AUC
- 10 IoU（Intersection-over-Union）指标
- 11 Top1与TopK