

模型评估

8.1 模型拟合统计

8.2 统计指标详解

8.3 调整单独采样

8.4 收益矩阵

样本内检验

样本内检验：使用训练集同期的数据



样本外检验：使用下一期的滚动数据



对评估数据集进行处理

- 评估数据集同样需要进行数据清洗、缺失值填补、分类变量WOE转换等操作；
- 在缺失值填补等操作时，需要使用使用训练数据集的统计量，而不是验证数据集的统计量。
- 如训练集变量用Income的均值400作为填补缺失值，测试集应用400填充，而不是重新计算均值

评估指标汇总

预测类型	统计量
决策 (Decisions)	准确率/误分类 利润/成本
排序 (Rankings)	ROC 指标 (一致性) Gini 指数 K-S统计量 提升度

决策类模型评估

该类模型的需求是回答“是不是？”。比如判别持身份证办业务的人是否为证件所有者。

混淆矩阵： 每给定一个阈值，就可以做出一个混淆矩阵		打分值		合计
		反应（预测=1）	未反应（预测=0）	
真实结果	呈现信号（真实=1）	A（击中） True Positive	B（漏报） False Negative	A + B
	未呈现信号（真实=0）	C（虚报） False Positive	D（正确否定） True Negative	C + D
合计		A + C	B + D	A + B + C + D

1. 正确率 = $(A+D)/(A+B+C+D)$
2. 灵敏度（**Sensitivity**；覆盖率recall）= $A/(A+B)$
3. 命中率(Precision、PV+)= $A/(A+C)$
4. 特异度 (**Specificity**；负例的覆盖率)= $D/(C+D)$
5. 负命中率(PV-) = $D/(D+B)$

评估指标汇总

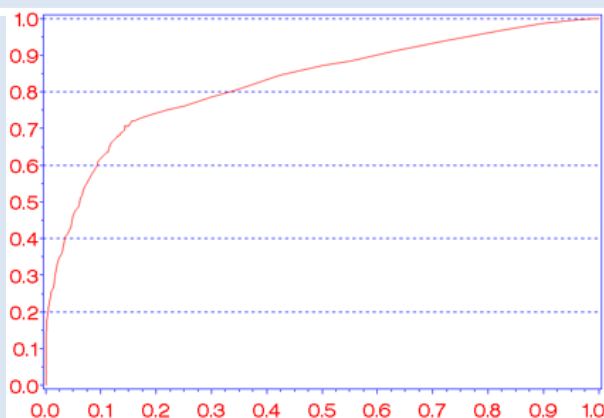
预测类型	统计量
决策 (Decisions)	精确性/误分类 利润/成本
排序 (Rankings)	ROC 指标 (一致性) Gini 指数 K-S统计量 提升度

排序类模型的评估指标

该类模型的需求是回答“会不会？”。比如预测一下客户违约的概率、营销响应的概率

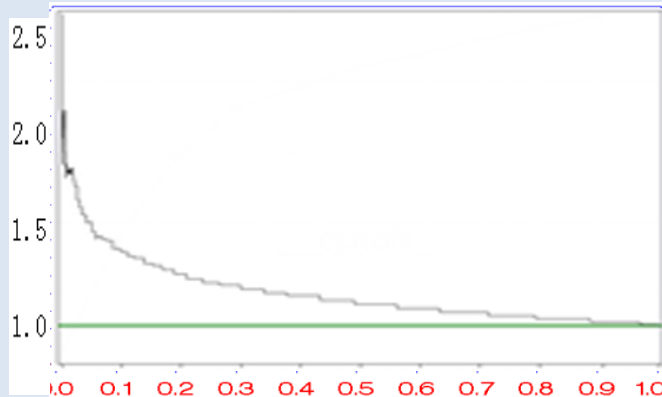
ROC曲线：用来描述模型分辨能力对角线以上的图形越高模型越好

X:1-特异度
Y: 灵敏度



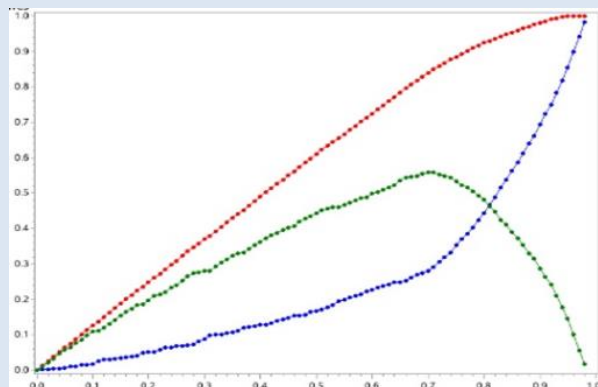
累积提升曲线：由于展示使用模型预测结果与随机情况下获取显性样本的

X:深度
Y: 正例的
累积密度
除以基准
概率



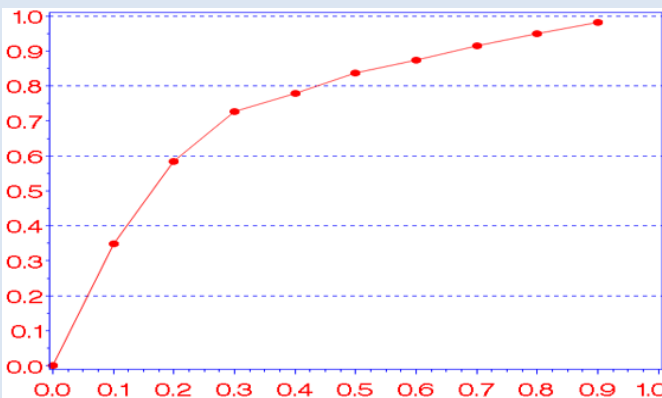
K-S曲线：用来描述模型对违约客户的分辨能力

X:深度
Y红：正例的
累积密度
Y蓝：负例的
累积密度
Y率：K-S值



洛伦兹曲线：用来描述预期违约客户的分布

X:深度
Y: 正例的
累积密度



第8章： 模型评估

8.1 模型拟合统计

8.2 统计指标详解

8.3 调整单独采样

8.4 收益矩阵

混淆矩阵 (Confusion Matrix)

ROC (Receiver Operating Characterstic) 曲线——接收者操作特征曲线。
最早应用于雷达信号检测领域，用于区分信号与噪声。

信号检测论能告诉我们什么？

例如：听觉：是否听到某事物的存在。

信号检测论：在听觉感受性相同的情况下，判断标准不一样。

①冒进：每次出现不会“漏报”，感觉有就报告。

②保守：每次出现不会“虚报”，没有把握不会报告。

		打分值(Predicted Class)		合计
		反应 (预测=1)	未反应 (预测=0)	
真实结果 (Actual Class)	呈现信号 (真实=1)	A (击中, True Positive)	B (漏报, False Negative)	A + B, Actual Positive
	未呈现信号 (真实=0)	C (虚报, False Positive)	D (正确否定, True Negative)	C + D, Actual Negative
合计		A + C, Predicted Positive	B + D, Predicted Negative	A + B + C + D

灵敏度 = $A / (A + B)$

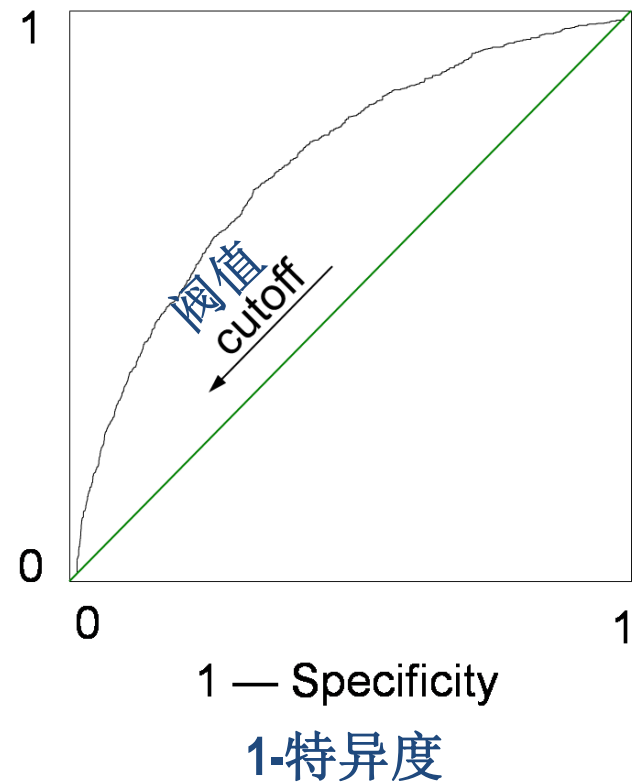
特异度 = $D / (C + D)$

ROC图形

	⑫ 阈值	⑫ 敏感度	⑫ 特异度
1	0.96	0.003	0.998
2	0.91	0.022	0.997
3	0.89	0.038	0.996
4	0.83	0.082	0.989
5	0.75	0.173	0.959
6	0.72	0.227	0.948
7	0.69	0.280	0.929
8	0.61	0.438	0.846
9	0.60	0.467	0.834
10	0.52	0.670	0.721
11	0.48	0.730	0.666
12	0.41	0.829	0.517
13	0.36	0.880	0.412
14	0.35	0.882	0.399
15	0.30	0.908	0.323
16	0.21	0.953	0.200
17	0.17	0.967	0.152
18	0.11	0.983	0.092
19	0.05	0.991	0.048
20	0.00	0.998	0.005

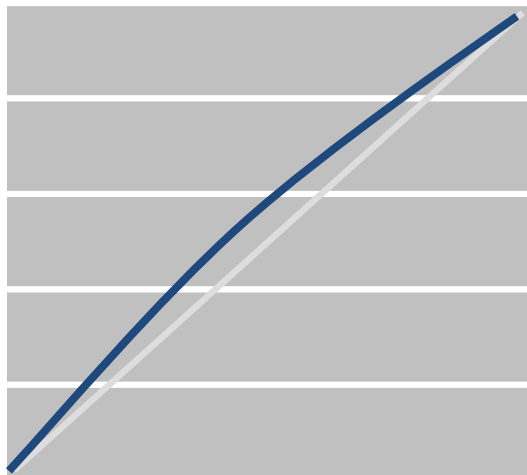
阈值下降

灵敏度
Sensitivity

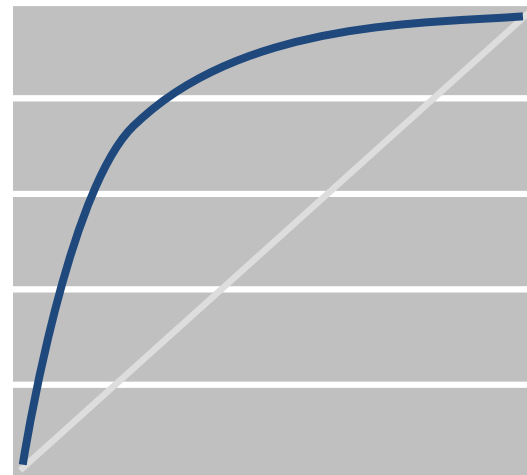
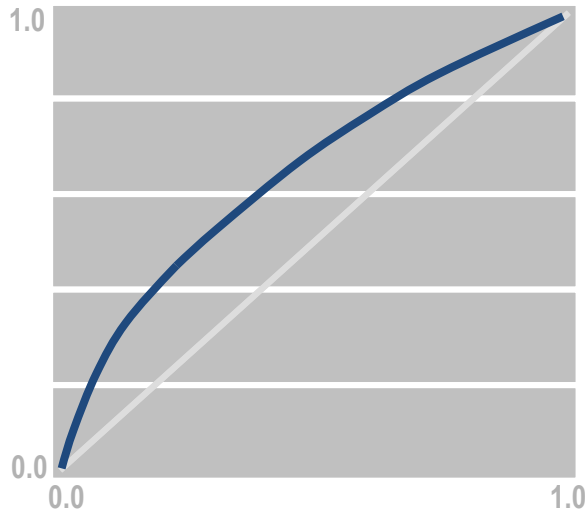


随着阈值的下降，灵敏度在升高，特异度在降低。

ROC 图形



弱的模型



强的模型

ROC曲线结果的取值在 $[0.5, 1]$ 。

一般来说，

$[0.5, 0.7)$ 表示效果较低；

$[0.7, 0.85)$ 表示效果一般；

$[0.85, 0.95)$ 表示效果良好；

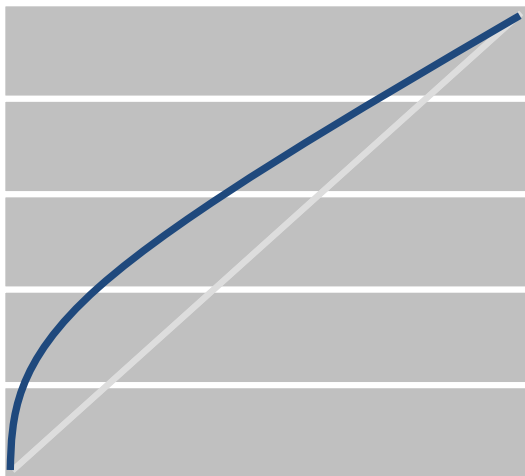
$[0.95, 1]$ 社会科学建模中不大可能出现。

注意：

①有时ROC曲线可能会落入对角线以下，这时需检查检验方向与状态值的对应关系

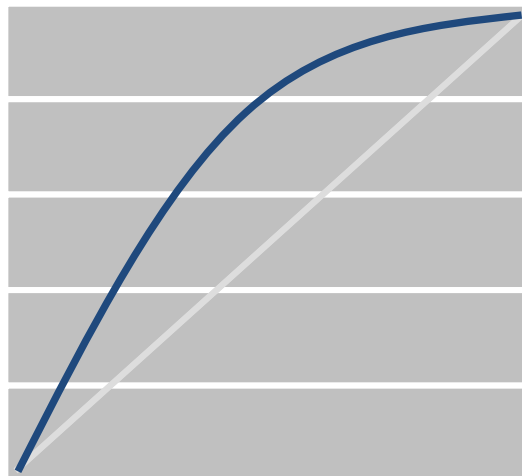
②如果某ROC曲线在对角线两边均有分布，需检查数据或专业背景。

ROC 图形



违约分值高处敏感

该模型在违约风险**高**人群中的预测能力较强，而在违约率**低**的部分较弱。有些业务需要做出这样的模型，比如汽车金融公司，业务需要只把违约风险非常高的客户筛选出来，而大部分客户授予分期付款。



违约分值低处敏感

该模型在违约风险**低**人群中的预测能力较强，而在违约率**高**的部分较弱。有些业务需要做出这样的模型，比如VIP信用卡产品，业务需要低风险客户较高的信用额度，因此需要明确哪些客户的违约风险很低。

如果使用过抽样

		预测结果		
		0	1	
实际	0	29	21	50
	1	17	33	50
		46	54	

样本

		预测结果		
		0	1	
实际	0	56	41	97
	1	1	2	3
		57	43	

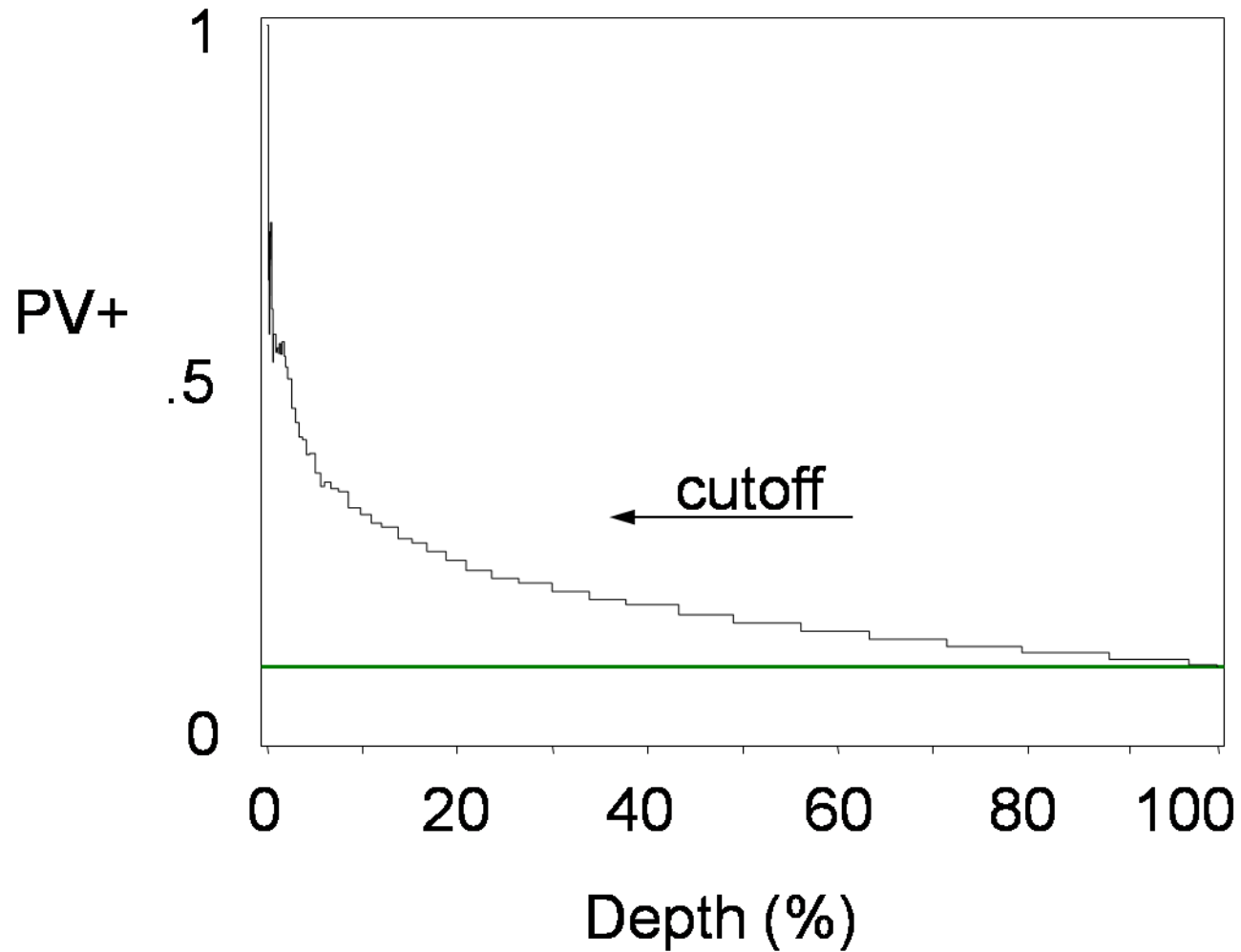
总体

过抽样调整

预测的分类

		预测的分类		
		0	1	
实际的分类	0	$\pi_0 \cdot Sp$	$\pi_0(1 - Sp)$	π_0
	1	$\pi_1(1 - Se)$	$\pi_1 \cdot Se$	π_1

Gains Chart



第8章： 模型评估

8.1 模型拟合统计

8.2 统计指标详解

8.3 调整单独采样

8.4 收益矩阵

评估指标汇总

预测类型	统计量
决策 (Decisions)	精确性/误分类 利润/成本
排序 (Rankings)	ROC 指标 (一致性) Gini 指数 K-S统计量
估计 (Estimates)	误差平方均值 SBC/可能性

对评估数据集进行处理

- 评估数据集同样需要进行数据清洗、缺失值填补、分类变量WOE转换等操作；
- 在缺失值填补等操作时，需要使用使用训练数据集的统计量，而不是验证数据集的统计量。

决策模型主要指标

- 精确性
- 误分类
- F统计量

混淆矩阵

预测 \ 实际	1	0
1	d True Positive	c False Negative
0	b False Positive	a True Negative

1. 准确率或者误分类率

$$\text{正确率} = (a+d)/(a+b+c+d)$$

2. 覆盖率 (**recall, Sensitivity**) $d/(c+d)$

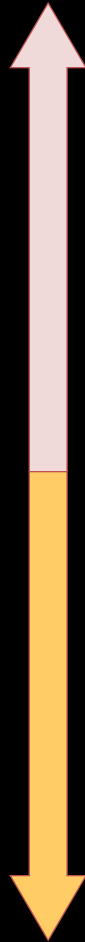
3. 命中率 (**Precision, PV+**) $d/(b+d)$

4. 负例的覆盖率 (**Specificity**) $a/(a+b)$

5. PV- $a/(a+c)$

混淆矩阵 (Confusion Matrix)

Predicted Probability	True Class
0.90	1
0.80	1
0.70	0
0.60	1
0.55	1
0.54	1
0.53	1
0.52	0
0.51	1
0.51	1
0.40	1
0.39	0
0.38	1
0.37	0
0.36	0
0.35	0
0.34	1
0.33	0
0.30	0
0.10	0



Confusion Matrix		True Results	
		0	1
Predicted Results	0	7	3
	1	2	8

Accuracy = 0.75 & Misclassification = 0.25

Precision = 0.8 & Recall = $\frac{8}{11} = 0.73$

F-Measure = 0.76

Sensitivity = True Positive Rate = Recall

Specificity = True Negative Rate = $\frac{7}{9} = 0.78$

混淆矩阵 (Confusion Matrix)

Confusion Matrix		True Results	
		0(Secondary)	1(Primary)
Predicted Results	0(Secondary)	TN(真陰性)	FN(偽陰性)
	1(Primary)	FP(偽陽性)	TP(真陽性)

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

$$\text{Precision} = \frac{\text{TP}}{\text{FP} + \text{TP}} \quad \text{Recall} = \frac{\text{TP}}{\text{FN} + \text{TP}}$$

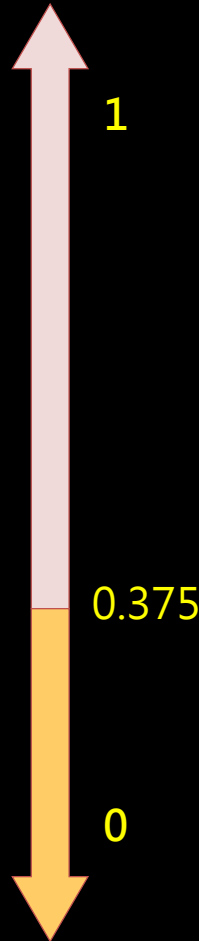
$$\text{F-Measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{Sensitivity} = \text{True Positive Rate} = \text{Recall}$$

$$\text{Specificity} = \text{True Negative Rate} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

混淆矩阵 (Confusion Matrix)

Predicted Probability	True Class
0.90	1
0.80	1
0.70	0
0.60	1
0.55	1
0.54	1
0.53	1
0.52	0
0.51	1
0.51	1
0.40	1
0.39	0
0.38	1
0.37	0
0.36	0
0.35	0
0.34	1
0.33	0
0.30	0
0.10	0



Predicted Results

Confusion Matrix		True Results	
		0	1
Predicted Results	0	6	1
	1	3	10

Accuracy = 0.8 & Misclassification = 0.2

Precision = 0.77 & Recall = $\frac{10}{11} = 0.91$

0.375 F-Measure = 0.83

Sensitivity = True Positive Rate = Recall

Specificity = True Negative Rate = $\frac{6}{9} = 0.67$

0

排序模型主要指标

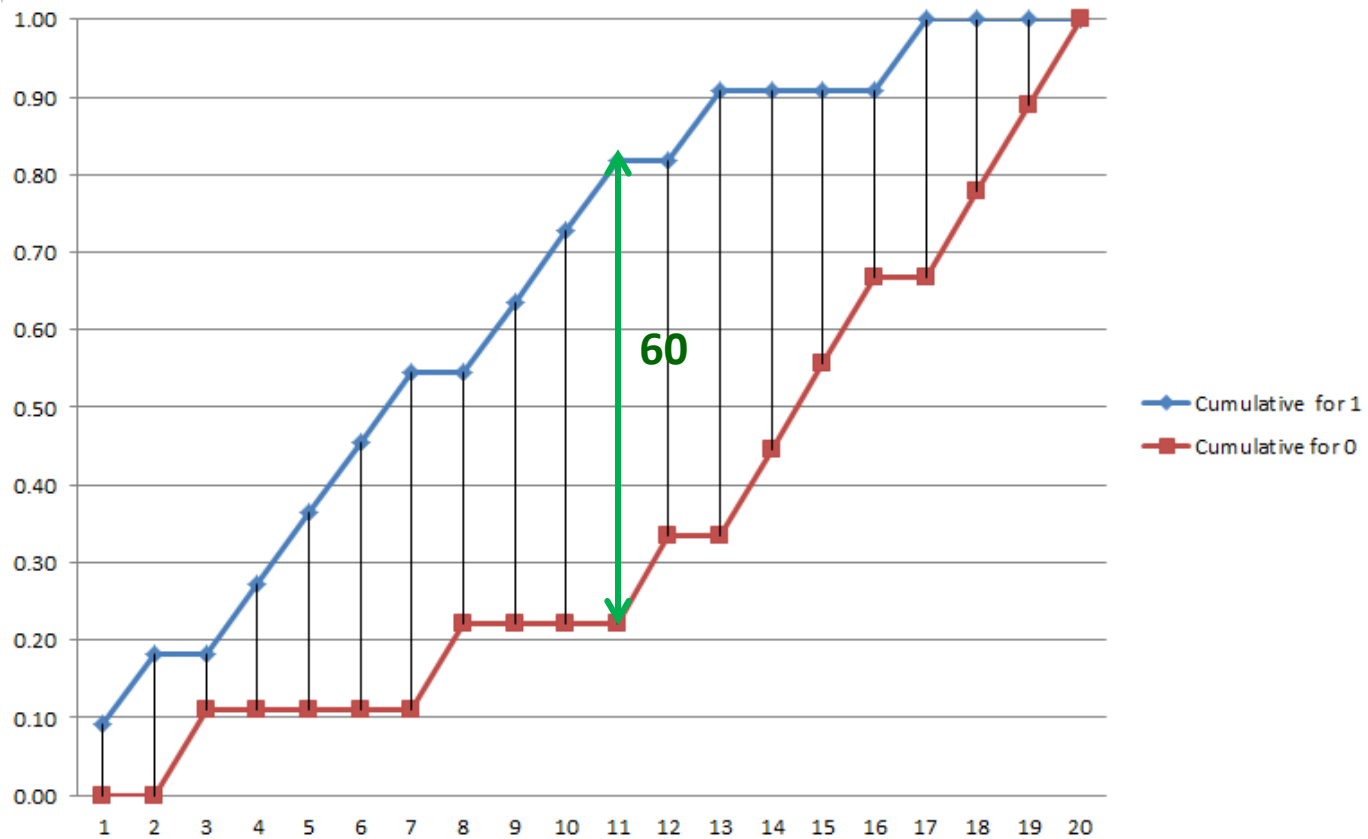
- K-S
- ROC
- Gini

K-S统计量

Predicted Probability	True Class	Cumulative for 1	Cumulative for 0	Difference*100
0.90	1	0.09	0.00	9
0.80	1	0.18	0.00	18
0.70	0	0.18	0.11	7
0.60	1	0.27	0.11	16
0.55	1	0.36	0.11	25
0.54	1	0.45	0.11	34
0.53	1	0.55	0.11	44
0.52	0	0.55	0.22	33
0.51	1	0.64	0.22	42
0.51	1	0.73	0.22	51
0.40	1	0.82	0.22	60
0.39	0	0.82	0.33	49
0.38	1	0.91	0.33	58
0.37	0	0.91	0.44	47
0.36	0	0.91	0.56	35
0.35	0	0.91	0.67	24
0.34	1	1.00	0.67	33
0.33	0	1.00	0.78	22
0.30	0	1.00	0.89	11
0.10	0	1.00	1.00	0

(K-S = 60)

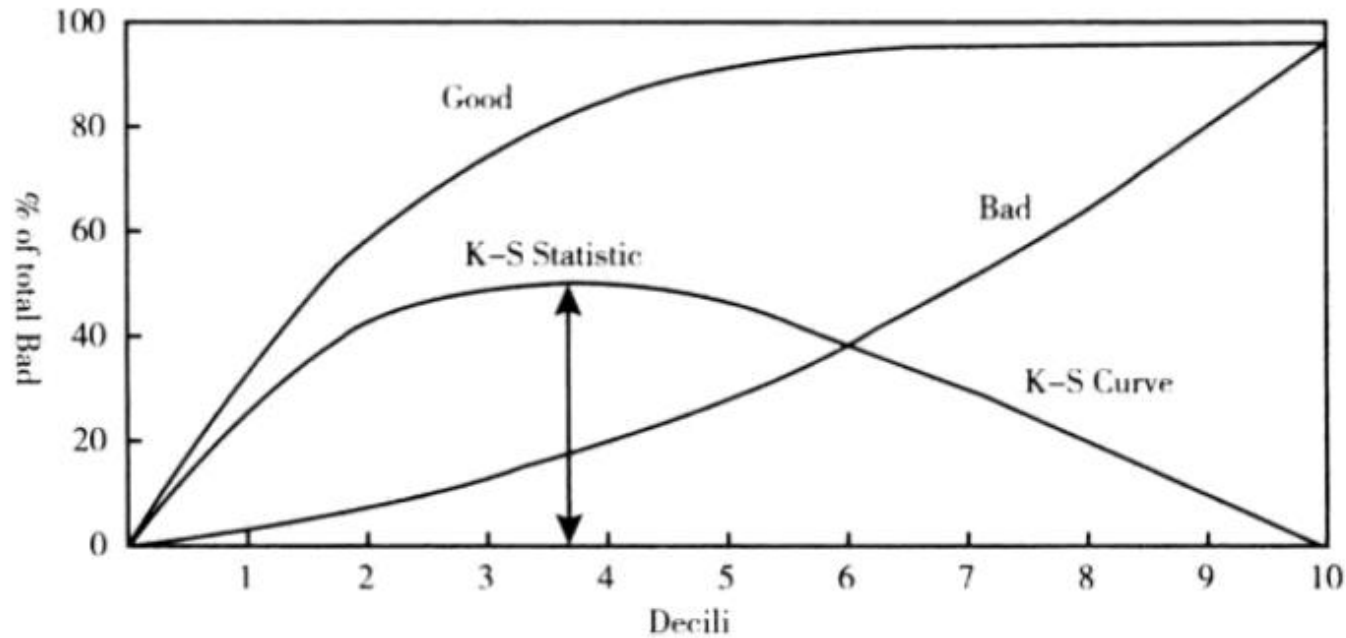
K-S Chart



K-S Chart

- K-S (Kolmogorov-Smirnov) chart measures performance of classification models
- K-S is a measure of the degree of separation between the positive and negative distributions
 - The K-S is 100 if the scores partition the population into two separate groups in which one group contains all the positives and the other all the negatives
 - If the model cannot differentiate between positives and negatives, then the K-S would be 0 (the model selects cases randomly)
 - The higher the value the better the model is at separating the positive from negative cases

K-S曲线



• K-S Statistics

- 小於20
 - 此模型無鑑別力
- 20~40之間
 - 此模型勉強接受
- 41~50之間
 - 此模型有好的區別能力

• K-S Statistics

- 51~60之間
 - 此模型有很好的區別能力
- 61~75之間
 - 此模型有非常好的區別能力
- 大於75
 - 此模型異常，可能有問題

Statistical Graphics – ROC Chart

Predicted Probability	True Class	Sensitivity	Specificity	1-Specificity
0.90	1	0.09	1.00	0.00
0.80	1	0.18	1.00	0.00
0.70	0	0.18	0.89	0.11
0.60	1	0.27	0.89	0.11
0.55	1	0.36	0.89	0.11
0.54	1	0.45	0.89	0.11
0.53	1	0.55	0.89	0.11
0.52	0	0.55	0.78	0.22
0.51	1	0.64	0.78	0.22
0.51	1	0.73	0.78	0.22
0.40	1	0.82	0.78	0.22
0.39	0	0.82	0.67	0.33
0.38	1	0.91	0.67	0.33
0.37	0	0.91	0.56	0.44
0.36	0	0.91	0.44	0.56
0.35	0	0.91	0.33	0.67
0.34	1	1.00	0.33	0.67
0.33	0	1.00	0.22	0.78
0.30	0	1.00	0.11	0.89
0.10	0	1.00	0.00	1.00

CM		True	
		0	1
P	0	8	8
	1	1	3

Accuracy = 0.55

Precision = 0.75

Recall = 0.27

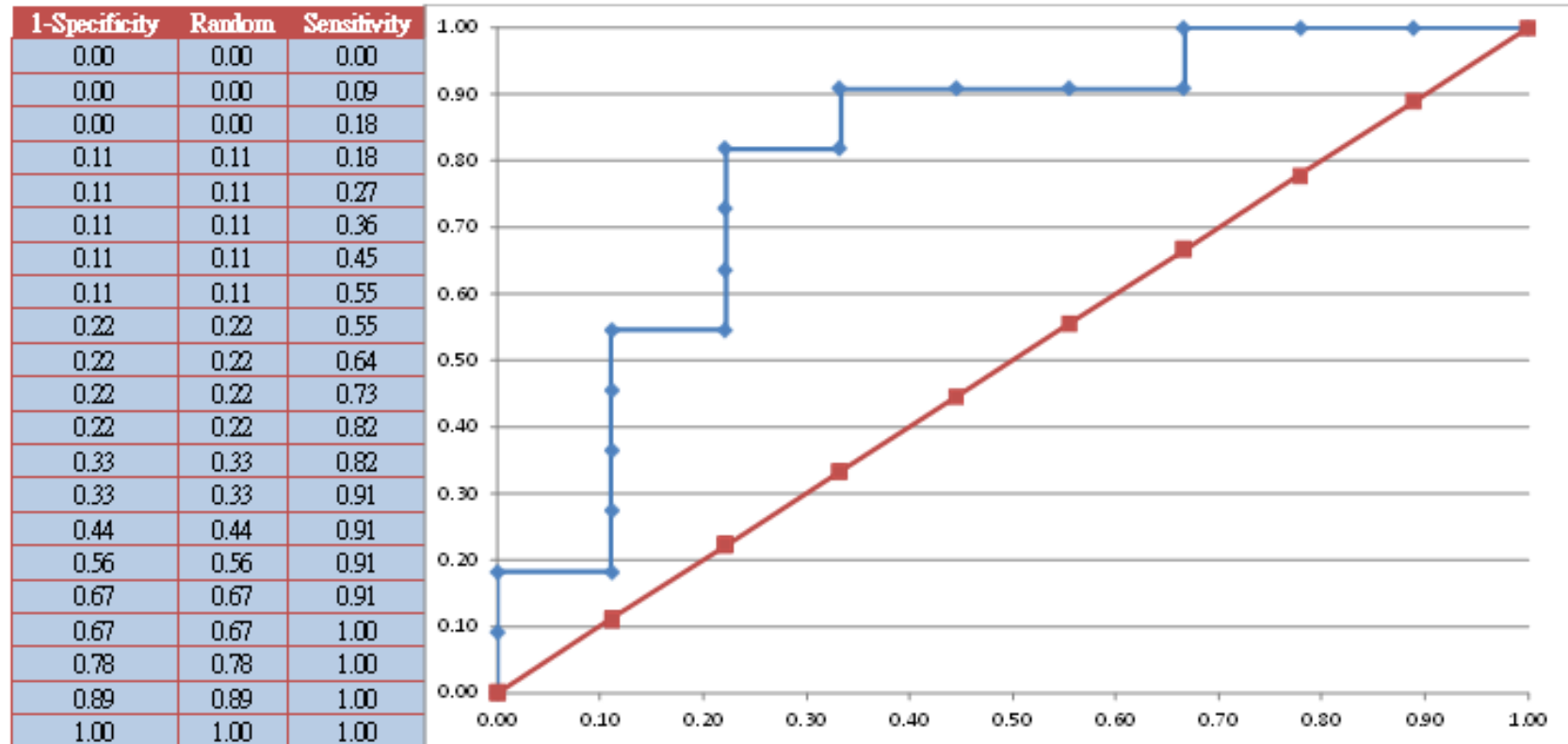
Sensitivity = Recall

Specificity
= True Negative Rate
= 0.89

Statistical Graphics – ROC Chart

Predicted Probability	True Class	Cumulative for 1	Cumulative for 0	Sensitivity	Specificity	1-Specificity
0.90	1	0.09	0.00	0.09	1.00	0.00
0.80	1	0.18	0.00	0.18	1.00	0.00
0.70	0	0.18	0.11	0.18	0.89	0.11
0.60	1	0.27	0.11	0.27	0.89	0.11
0.55	1	0.36	0.11	0.36	0.89	0.11
0.54	1	0.45	0.11	0.45	0.89	0.11
0.53	1	0.55	0.11	0.55	0.89	0.11
0.52	0	0.55	0.22	0.55	0.78	0.22
0.51	1	0.64	0.22	0.64	0.78	0.22
0.51	1	0.73	0.22	0.73	0.78	0.22
0.40	1	0.82	0.22	0.82	0.78	0.22
0.39	0	0.82	0.33	0.82	0.67	0.33
0.38	1	0.91	0.33	0.91	0.67	0.33
0.37	0	0.91	0.44	0.91	0.56	0.44
0.36	0	0.91	0.56	0.91	0.44	0.56
0.35	0	0.91	0.67	0.91	0.33	0.67
0.34	1	1.00	0.67	1.00	0.33	0.67
0.33	0	1.00	0.78	1.00	0.22	0.78
0.30	0	1.00	0.89	1.00	0.11	0.89
0.10	0	1.00	1.00	1.00	0.00	1.00

Statistical Graphics – ROC Chart



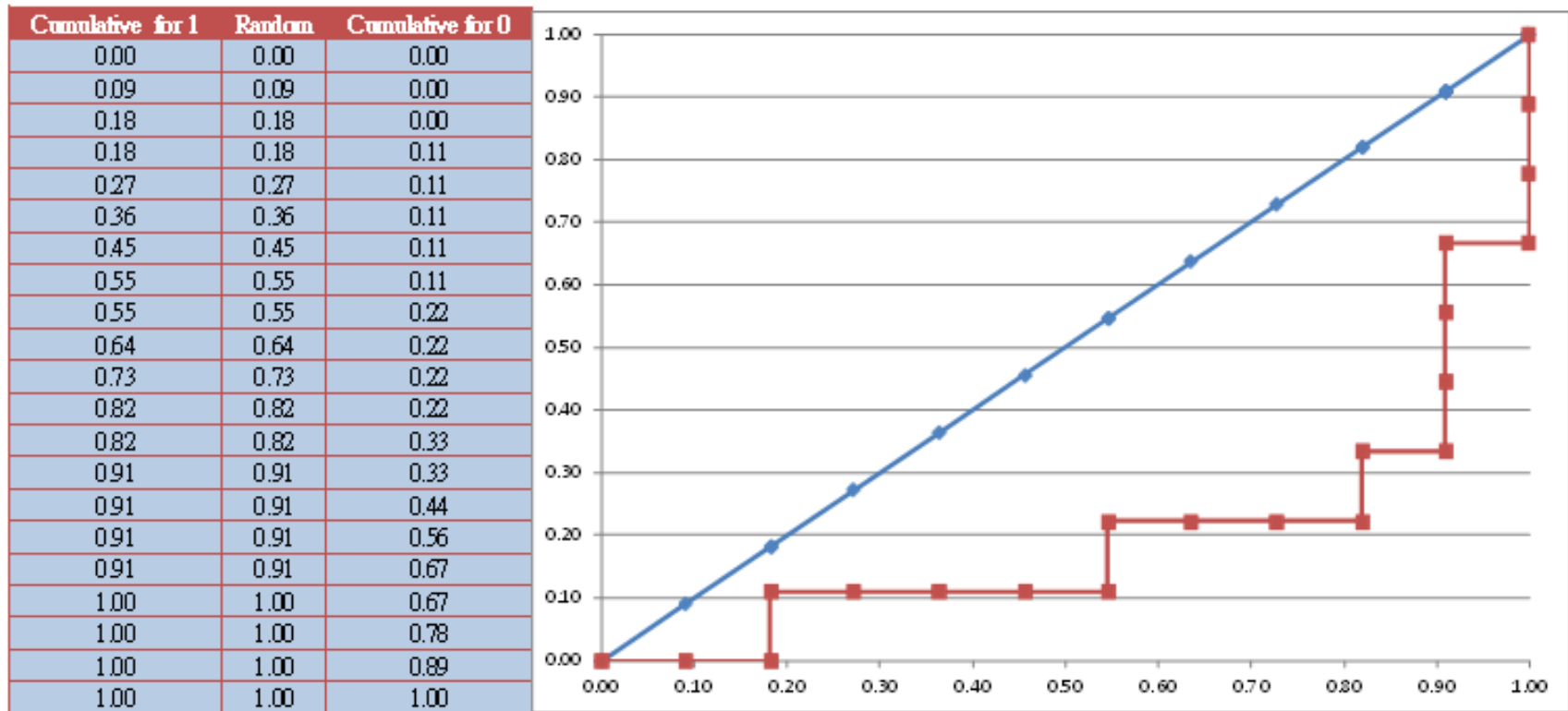
Area Under the Curve (AUC)

- Area under ROC curve is often used as a measure of quality of the classification models
- A random classifier has an area under the curve of 0.5, while AUC for a perfect classifier is equal to 1
- In practice, most of the classification models have an AUC between 0.5 and 1

Gini Coefficient

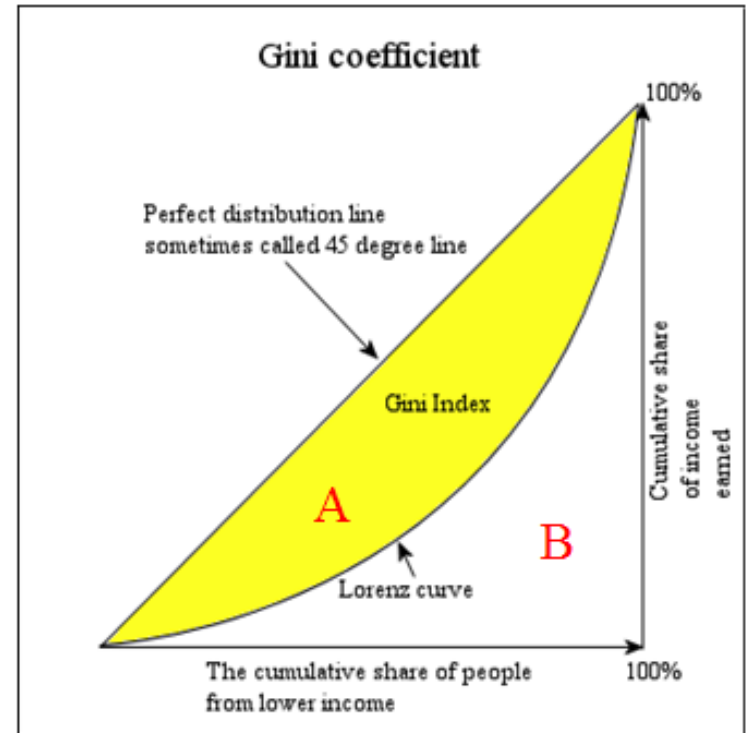
Predicted Probability	True Class	Cumulative for 1	Cumulative for 0	X	Y
0.90	1	0.09	0.00	0.09	0.00
0.80	1	0.18	0.00	0.18	0.00
0.70	0	0.18	0.11	0.18	0.11
0.60	1	0.27	0.11	0.27	0.11
0.55	1	0.36	0.11	0.36	0.11
0.54	1	0.45	0.11	0.45	0.11
0.53	1	0.55	0.11	0.55	0.11
0.52	0	0.55	0.22	0.55	0.22
0.51	1	0.64	0.22	0.64	0.22
0.51	1	0.73	0.22	0.73	0.22
0.40	1	0.82	0.22	0.82	0.22
0.39	0	0.82	0.33	0.82	0.33
0.38	1	0.91	0.33	0.91	0.33
0.37	0	0.91	0.44	0.91	0.44
0.36	0	0.91	0.56	0.91	0.56
0.35	0	0.91	0.67	0.91	0.67
0.34	1	1.00	0.67	1.00	0.67
0.33	0	1.00	0.78	1.00	0.78
0.30	0	1.00	0.89	1.00	0.89
0.10	0	1.00	1.00	1.00	1.00

Gini Coefficient



Gini Coefficient vs. ROC Index

- The Gini coefficient is calculated as a ratio of the areas on the **Lorenz curve** diagram
 - A : the area between the line of perfect equality and Lorenz curve
 - B : the area underneath the Lorenz curve
 - Gini coefficient is $A/(A+B)$



$$\text{ROC Index} = A + 0.5$$

$$\text{Gini Coefficient} = A / (A+B) = 2A = 2 * (\text{ROC Index} - 0.5)$$

第8章： 模型评估

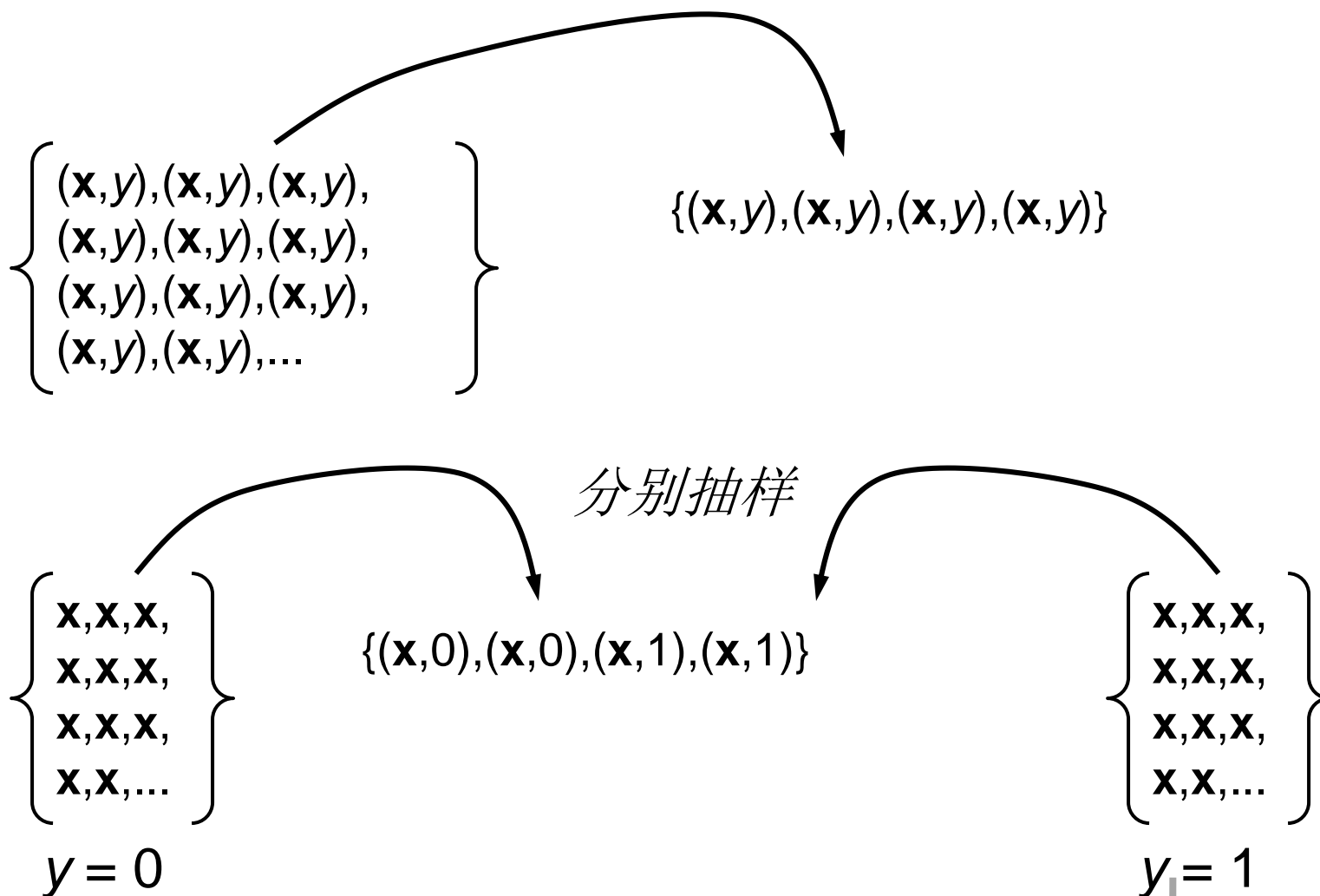
8.1 模型拟合统计

8.2 统计指标详解

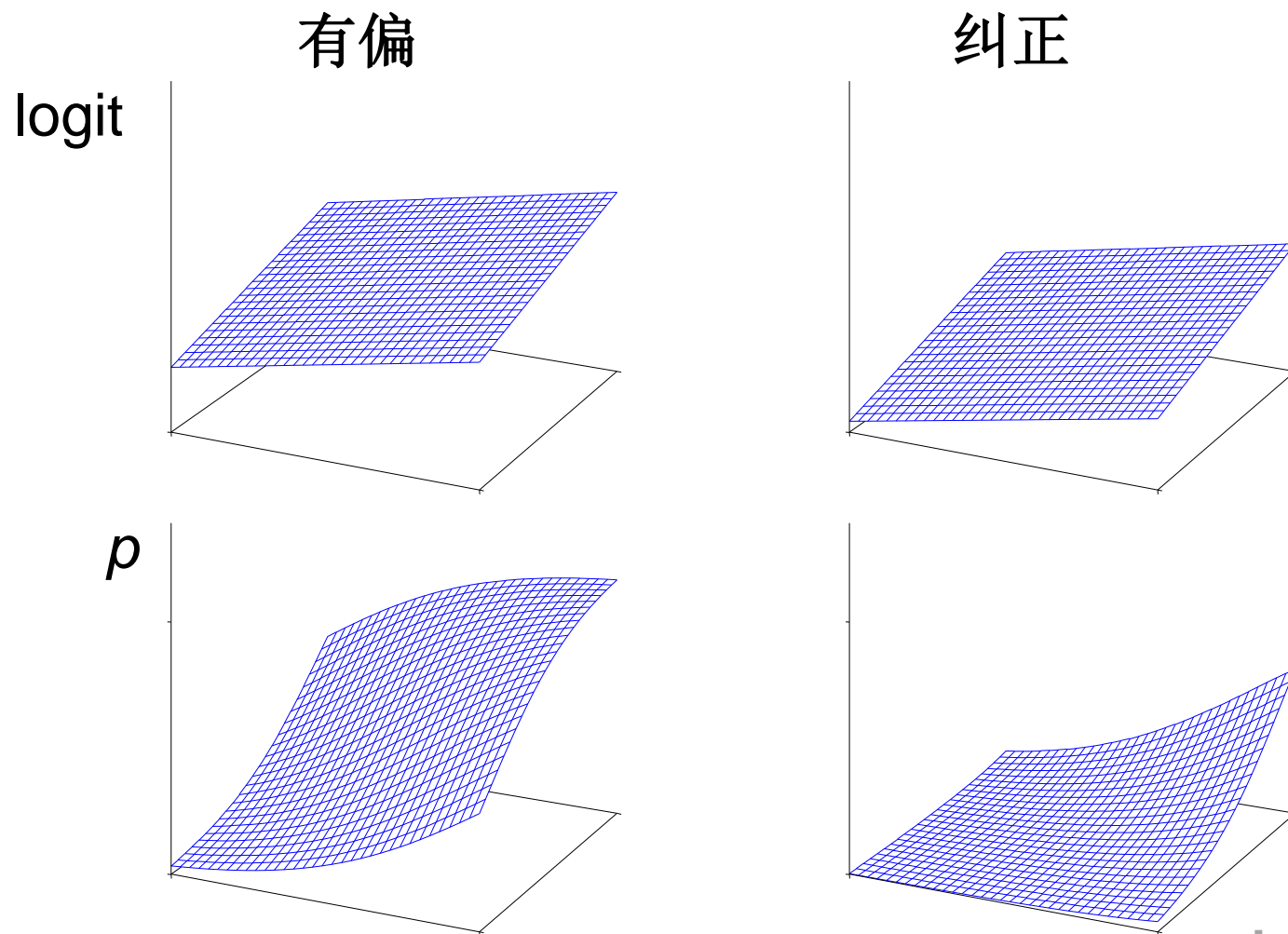
8.3 调整单独采样

8.4 收益矩阵

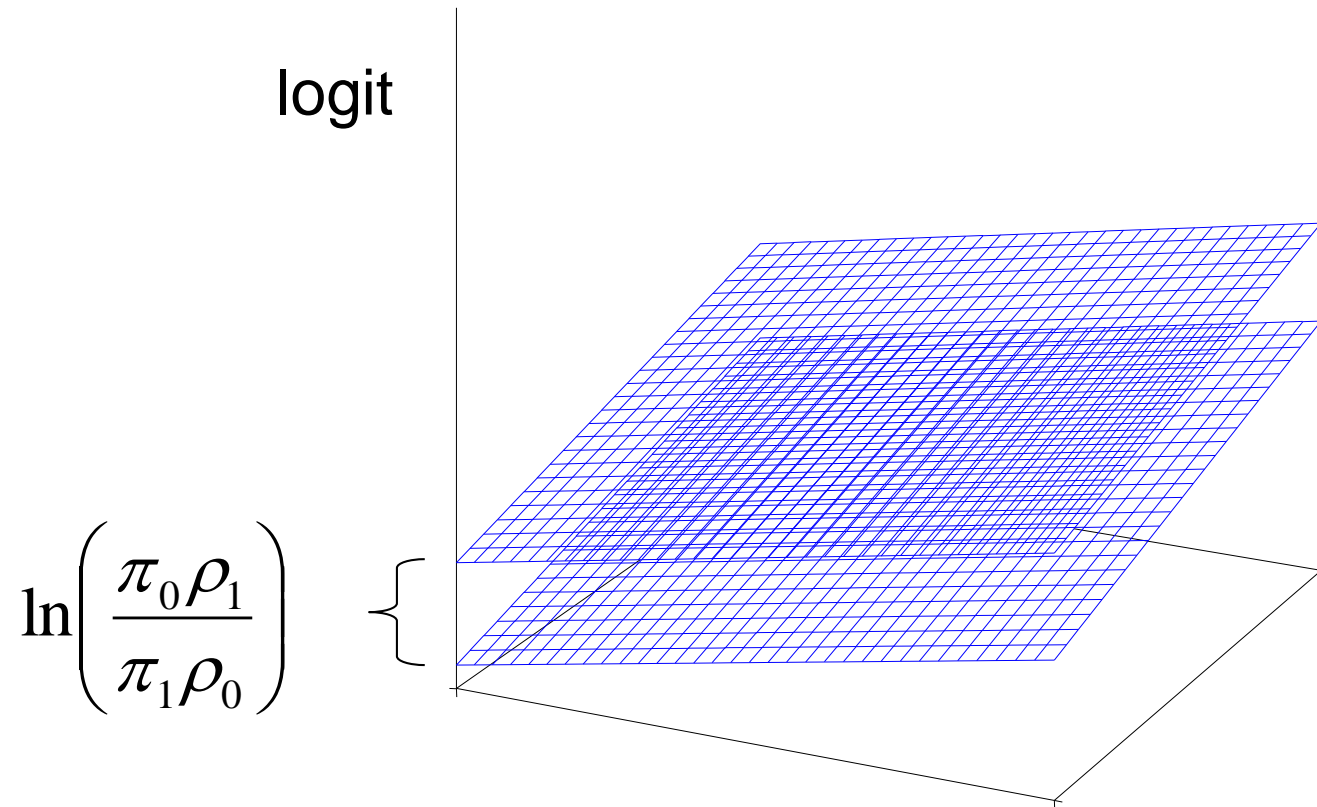
抽样设计



过度抽样的影响



预测值平移



第8章： 模型评估

8.1 模型拟合统计

8.2 统计指标详解

8.3 调整单独采样

8.4 收益矩阵

利润矩阵

实际 \ 预测	1	0
1	50	0
0	-20	0


$$P \cdot 50 + (1-P)(-20) > P \cdot 0 + (1-P) \cdot 0$$

$$70p - 20 > 0$$

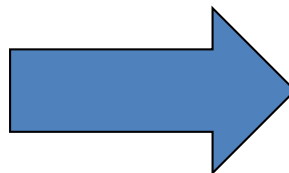
$$P > 2/7$$

成本矩阵

<div> <div>预测</div> <div>实际</div> </div>	1	0
1	0	5
0	1	0

$$(1 - p) \text{cost}_{FP} < p * \text{cost}_{FN}$$

$$p > \frac{1}{1 + \frac{\text{cost}_{FN}}{\text{cost}_{FP}}}$$



$$p > \frac{1}{1 + \frac{5}{1}} = \frac{1}{6}$$