

2020.2.18 - 2020.6.2 9:50-12:15@6C102雨课堂+腾讯会议
《模式识别与机器学习》



第三章 学习机器性能评估

2020.2.25



Xuegong Zhang



怎样评价一个学习机器的好坏？



机器学习的基本要素

- 怎样造一个学习机器？
 - 它需要老师
 - 我们设计它（特征和模型）
 - 它需要训练/学习材料
 - 训练数据
 - 我们需要为它树立学习的目标
 - 目标函数、学习准则
 - 我们需要告诉它怎样学
 - 学习训练算法



Xuegong Zhang

2





怎样评价一个学习机器的好坏？



机器学习的基本要素

• 怎样造一个学习机器？

- 它需要老师
→ 我们设计它（特征和模型）
- 它需要训练/学习材料
→ 训练数据
- 我们需要为它树立学习的目标
→ 目标函数、学习准则
- 我们需要告诉它怎样学
→ 学习训练算法



• 学习机器的“价值观”

- 目标函数

- 零错误
- 最少错误
- 最大似然

都是在训练数据上

- 每个机器都觉得自己是最好的，我们应该怎么看？

Xuegong Zhang

3



3.1 分类器的评价 ——错误率的定义和计算

Xuegong Zhang

4



问题1：怎样计算错误？
问题2：所有错误都一样吗？

Xuegong Zhang

5



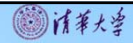
错误率 (Error Rate)

- **错误率**：样本集中分类错误样本所占比例，有时简称为“错误”
- **训练错误率**：训练样本集上的错误率
 - 同义词：视在错误 (apparent error)、重带入错误 (re-substitution error)、经验错误 (empirical error)、...

Xuegong Zhang

6

错误率 (Error Rate)



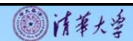
- **错误率**：样本集中分类错误样本所占比例，有时简称为“错误”
- **训练错误率**：训练样本集上的错误率
 - 同义词：视在错误 (apparent error)、重带入错误 (re-substitution error)、经验错误 (empirical error)、...
- **独立测试集**（简称测试集）：训练集之外独立同分布的样本 (设样本数 N)
- **测试错误率**：测试集上的错误率

$$\hat{\varepsilon} = \frac{k}{N} \quad k \text{ 是测试集上分类错误样本的数目}$$

Xuegong Zhang

7

错误率 (Error Rate)



- **错误率**：样本集中分类错误样本所占比例，有时简称为“错误”
- **训练错误率**：训练样本集上的错误率
 - 同义词：视在错误 (apparent error)、重带入错误 (re-substitution error)、经验错误 (empirical error)、...
- **独立测试集**（简称测试集）：训练集之外独立同分布的样本 (设样本数 N)
- **测试错误率**：测试集上的错误率

$$\hat{\varepsilon} = \frac{k}{N} \quad k \text{ 是测试集上分类错误样本的数目}$$

- **真实错误率**：在未来未知的样本集上的错误率，也称作**期望错误率**
 - 对分类器性能的真实评价，但无法预知
 - 训练错误率、测试错误率都可以作为对真实错误率的估计
 - 训练错误率易偏乐观

Xuegong Zhang

8

真实错误率的估计



- 测试错误率:

$$\hat{\varepsilon} = k/N$$



是否是对真实错误率 ε 的准确估计?

Xuegong Zhang

9

单选题 1分

设置

测试错误率是否是对真实错误率的可靠估计?

- ☐ A 是, 只要测试集与未来样本独立同分布
- ☐ B 否, 即使测试集与未来样本集独立同分布
- ☐ C 是, 只要测试集与训练样本集独立同分布
- ☒ D 不一定, 看情况

提交

10

多选题 1分

设置

测试错误率作为真实错误率的可靠估计是有条件的，最重要的条件是什么？

- ☒ A 测试集样本与训练集样本独立
- ☒ B 测试集样本与未来样本独立同分布
- ☒ C 测试集样本数足够大
- ☐ D 训练集样本数足够大

提交

11

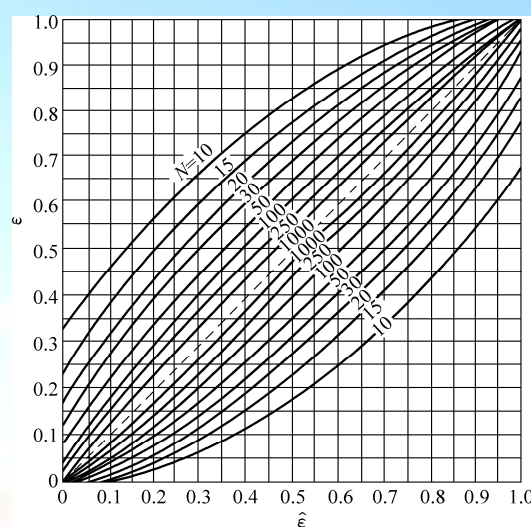
真实错误率的估计



- 测试错误率：

$$\hat{\varepsilon} = k/N$$

是否是对真实错误率 ε 的准确估计？



Xuegong Zhang

12

真实错误率的估计



- 测试错误率:

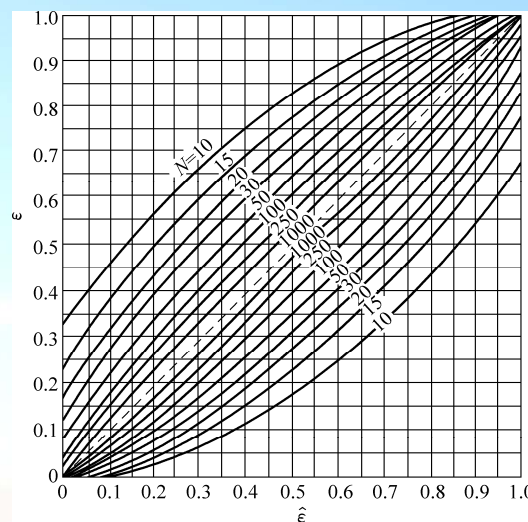
$$\hat{\varepsilon} = k/N$$

是否是对真实错误率 ε 的准确估计?

重要结论:

测试错误率是对真实错误率的无偏估计, 但估计的置信范围取决于测试样本数目, 有可能会很大。

→ 为得到准确可靠的估计, 需要大量测试样本



Xuegong Zhang

13

样本数的矛盾



- 样本获取通常是受各种因素制约的
 - 为了可靠评估分类器性能, 我们需要尽可能多测试样本
 - 但我们希望尽可能多的样本用于训练
- 矛盾:
 - 更多训练样本 → 更好的分类器性能
 - 更小的测试集 → 对分类器性能的评估不可靠
 - 更多测试样本 → 更准确的分类器性能估计
 - 更小的训练集 → 分类器性能更差



Xuegong Zhang

14

样本数的矛盾



- 样本获取通常是受各种因素制约的
 - 为了可靠评估分类器性能，我们需要尽可能多测试样本
 - 但我们希望尽可能的样本用于训练
- 矛盾：
 - 更多训练样本 → 更好的分类器性能
 - 更小的测试集 → 对分类器性能的评估不可靠
 - 更多测试样本 → 更准确的分类器性能估计
 - 更小的训练集 → 分类器性能更差



等会儿，为什么希望更多训练样本？

Xuegong Zhang

15

样本数的矛盾



- 样本获取通常是受各种因素制约的
 - 为了可靠评估分类器性能，我们需要尽可能多测试样本
 - 但我们希望尽可能的样本用于训练
- 矛盾：
 - 更多训练样本 → 更好的分类器性能 关于学习的理论
 - 更小的测试集 → 对分类器性能的评估不可靠
 - 更多测试样本 → 更准确的分类器性能估计
 - 更小的训练集 → 分类器性能更差 关于学习的理论



好吧，权且信之，怎么解决这个矛盾？

Xuegong Zhang

16

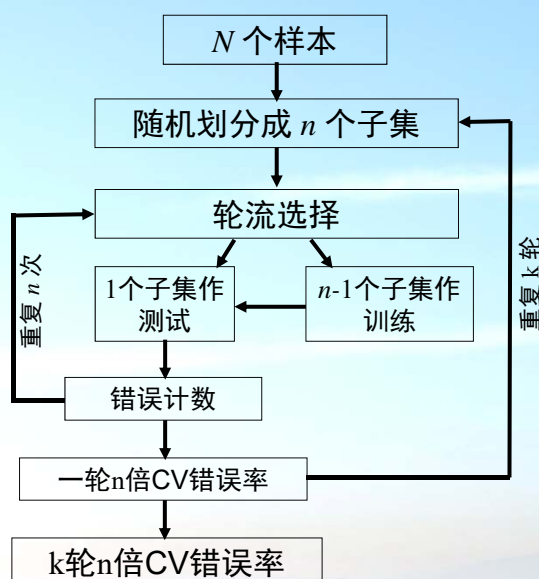
交叉验证 (Cross Validation (CV))



- 解决训练样本集与测试样本集矛盾的有效方案

– n 倍交叉验证 (n -fold cross validation)

- 几乎无偏估计
- 测试集大，估计方差小
- 略偏保守



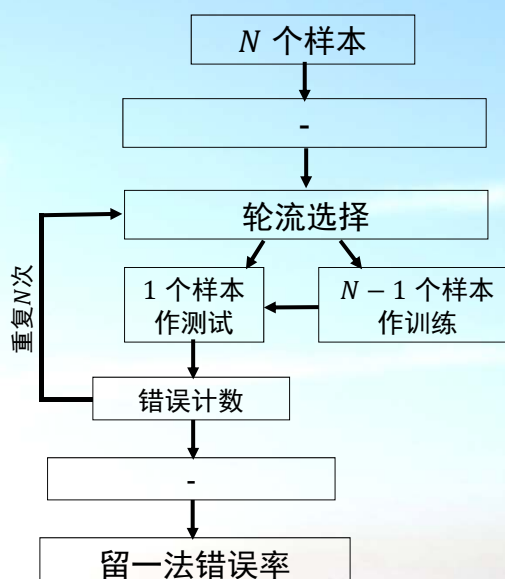
Xuegong Zhang

17

留一法交叉验证 (LOOCV)



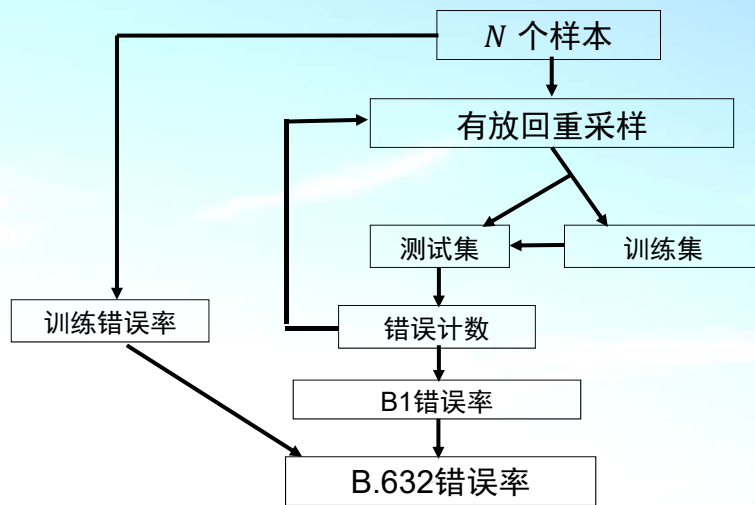
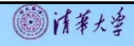
- 交叉验证的极端情况
- 样本少时被广泛应用



Xuegong Zhang

18

自举法与B.632估计



$$B.632+ = (1 - w) AE + w B1$$

$$B.632 = 0.368 * AE + 0.632 * B1$$

Xuegong Zhang

C. Ambrose & G.J. McLachlan, Selection bias in gene extraction on basis of microarray gene-expression data, *PNAS*, **99**(10): 6567-6572, 2002

19

休息1分钟



Xuegong Zhang

20



问题1：怎样计算错误？
问题2：所有错误都一样吗？

Xuegong Zhang

21

诊断测试中的重要术语



- **Positive 阳性**：有病（待检测疾病存在，如nCoV阳性）
- **Negative 阴性**：无病（待检测疾病不存在，如nCoV阴性）
- **Diagnostic test 诊断测试**：用某种技术检测疾病

	有病 D^+	无病 D^-
检测阳性 T^+	TP 真阳性	FP 假阳性
检测阴性 T^-	FN 假阴性	TN 真阴性

Xuegong Zhang

22

诊断测试中的重要术语



- **Positive 阳性**: 有病 (待检测疾病存在, 如nCoV阳性)
- **Negative 阴性**: 无病 (待检测疾病不存在, 如nCoV阴性)
- **Diagnostic test 诊断测试**: 用某种技术检测疾病
- **检测结果统计指标**:

- **Sensitivity 灵敏度**

$$P(T^+|D^+) = TP / (TP + FN)$$

- **Specificity 特异度**

$$P(T^-|D^-) = TN / (TN + FP)$$

- **Prevalence 患病率**: 人群中患病的概率, 亦称pre-test probability

$$D^+ / (D^+ + D^-)$$

- 检验的预测价值 (Predictive value)

- **Discovery rate 发现率**: $P(D^+|T^+) = TP / (TP + FP)$

- **False discovery rate 误发现率**: $P(D^-|T^+) = FP / (TP + FP)$

- **Accuracy 精度/准确率**: $(TP + TN) / (D^+ + D^-)$

	有病 D^+	无病 D^-
检测阳性 T^+	TP 真阳性	FP 假阳性
检测阴性 T^-	FN 假阴性	TN 真阴性

Xuegong Zhang

23

两类错误率



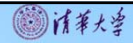
- **假设检验的术语**:
 - 空假设 H_0 与备择假设 H_1 (null hypothesis vs. alternative hypothesis)

统计决策	真实状态	
	H_0 真 (N_0) 阴性	H_1 真 (N_1) 阳性
拒绝 H_0 (检测阳性)	Type I error 第一类错误 假阳性 (FP) α (FP/N_0)	检测正确 真阳性 (TP) 灵敏度sensitivity (TP/N_1)
接受 H_0 (检测阴性)	检测正确 真阴性 (TN) 特异度specificity (TN/N_0)	Type II error 第二类错误 假阴性 (FN) β (FN/N_1)

Xuegong Zhang

24

两类错误率



- 假设检验的术语：
 - 空假设H0与备择假设H1 (null hypothesis vs. alternative hypothesis)

统计决策	真实状态	
	H0真 (N0) 阴性	H1真 (N1) 阳性
拒绝 H0 (检测阳性)	Type I error 第一类错误 假阳性 (FP) α (FP/N0)	检测正确 真阳性 (TP) 灵敏度sensitivity (TP/N1)
接受 H0 (检测阴性)	检测正确 真阴性 (TN) 特异度specificity (TN/N0)	Type II error 第二类错误 假阴性 (FN) β (FN/N1)

Sensitivity (power) = $TP / (TP + FN) = 1 - \beta$ a.k.a. detection rate, recall, ...

Specificity = $TN / (TN + FP) = 1 - \alpha$

Xuegong Zhang

25

More Jargons



Terminology and derivations from a confusion matrix	
condition positive (P) the number of real positive cases in the data	
condition negative (N) the number of real negative cases in the data	
true positive (TP) eqv. with hit	
true negative (TN) eqv. with correct rejection	
false positive (FP) eqv. with false alarm, Type I error	
false negative (FN) eqv. with miss, Type II error	
sensitivity, recall, hit rate, or true positive rate (TPR) $TPR = \frac{TP}{P} = \frac{TP}{TP + FN} = 1 - FNR$	miss rate or false negative rate (FNR) $FNR = \frac{FN}{P} = \frac{FN}{FN + TP} = 1 - TPR$
specificity, selectivity or true negative rate (TNR) $TNR = \frac{TN}{N} = \frac{TN}{TN + FP} = 1 - FPR$	fall-out or false positive rate (FPR) $FPR = \frac{FP}{N} = \frac{FP}{FP + TN} = 1 - TNR$
precision or positive predictive value (PPV) $PPV = \frac{TP}{TP + FP}$	false discovery rate (FDR) $FDR = \frac{FP}{FP + TP} = 1 - PPV$
negative predictive value (NPV) $NPV = \frac{TN}{TN + FN}$	false omission rate (FOR) $FOR = \frac{FN}{FN + TN} = 1 - NPV$
	accuracy (ACC) $ACC = \frac{TP + TN}{P + N} = \frac{TP + TN}{TP + TN + FP + FN}$
	F1 score is the harmonic mean of precision and sensitivity $F_1 = 2 \cdot \frac{PPV \cdot TPR}{PPV + TPR} = \frac{2TP}{2TP + FP + FN}$
	Matthews correlation coefficient (MCC) $MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$
	Informedness or Bookmaker Informedness (BM) $BM = TPR + TNR - 1$
	Markedness (MK) $MK = PPV + NPV - 1$
	Sources: Fawcett (2006), Powers (2011), and Ting (2011) [2] [3] [4]

Xuegong Zhang

Wikipedia

26

More Jargons

		True condition		Prevalence = $\frac{\sum \text{Condition positive}}{\sum \text{Total population}}$	Accuracy (ACC) = $\frac{\sum \text{True positive} + \sum \text{True negative}}{\sum \text{Total population}}$
		Condition positive	Condition negative		
Predicted condition	Predicted condition positive	True positive , Power	False positive , Type I error	Positive predictive value (PPV), Precision = $\frac{\sum \text{True positive}}{\sum \text{Predicted condition positive}}$	False discovery rate (FDR) = $\frac{\sum \text{False positive}}{\sum \text{Predicted condition positive}}$
	Predicted condition negative	False negative , Type II error	True negative	False omission rate (FOR) = $\frac{\sum \text{False negative}}{\sum \text{Predicted condition negative}}$	Negative predictive value (NPV) = $\frac{\sum \text{True negative}}{\sum \text{Predicted condition negative}}$
		True positive rate (TPR), Recall, Sensitivity, probability of detection = $\frac{\sum \text{True positive}}{\sum \text{Condition positive}}$	False positive rate (FPR), Fall-out, probability of false alarm = $\frac{\sum \text{False positive}}{\sum \text{Condition negative}}$	Positive likelihood ratio (LR+) = $\frac{\text{TPR}}{\text{FPR}}$	Diagnostic odds ratio (DOR) = $\frac{\text{LR+}}{\text{LR-}}$ F ₁ score = $\frac{2}{\frac{1}{\text{Recall}} + \frac{1}{\text{Precision}}}$
		False negative rate (FNR), Miss rate = $\frac{\sum \text{False negative}}{\sum \text{Condition positive}}$	Specificity (SPC), Selectivity, True negative rate (TNR) = $\frac{\sum \text{True negative}}{\sum \text{Condition negative}}$	Negative likelihood ratio (LR-) = $\frac{\text{FNR}}{\text{TNR}}$	

术语太乱搞不清怎么办？

- 在做实验和写文章时，把自己用的指标的定义明确写下来。

Wikipedia

Xuegong Zhang

27

15:34 0.00K/s 移动 4G HD 67%

×

华大癌变

原创：伯通 虎嗅网 今天

2017年9月3日，一个男婴降生在湖南省长沙市望城区妇幼保健院。不幸的是，他的出生伴随着“13号染色体长臂缺失综合症”、“脑发育不良”、“虹膜缺损”等一系列缺陷和疾病。这意味着，小男孩很可能会智力障碍、生长迟缓、外表异常，几乎无法正常长大。

这个在瞬间被拖入深渊的家庭，立刻向此前负责产检的医院及华大基因展开了诘问和斥责




例：他们是否说了谎？

15:34 0.00K/s 移动 4G HD 67%

×

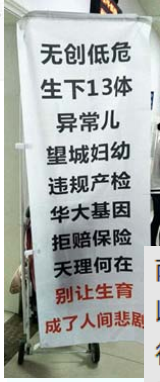
华大癌变

原创：伯通 虎嗅网 今天



2017年9月3日，一个男婴降生在湖南省长沙市望城区妇幼保健院。不幸的是，他的出生伴随着“13号染色体长臂缺失综合症”、“脑发育不良”、“虹膜缺损”等一系列缺陷和疾病。这意味着，小男孩很可能会智力障碍、生长迟缓、外表异常，几乎无法正常长大。

这个在瞬间被拖入深渊的家庭，立刻向此前负责产检的医院及华大基因展开了诘问和斥责



例：他们是否说了谎？

南方周末记者发现，早在2015年12月底，华大基因便以无创产前基因检测准确率高达99.9%作为宣传点，进行全面推广。但在2018年7月13日晚间，华大基因对外发布《深圳华大基因股份有限公司关于媒体报道的澄清公告》（下称《公告》）中这一数字悄悄发生了变化。


《公告》称，“国家卫计委发布《通知》中，对于常规染色体非整倍体的检出率做了以下要求：21 三体、18 三体、13 三体的检出率分别不低于 95%、85%和 70%，而非 100%。无创基因检测作为一项筛查技术存在假阴性的风险。”

15:34 0.00K/s 移动 4G HD 67%

×


华大癌变

原创：伯通 虎嗅网 今天



2017年9月3日，一个男婴降生在湖南省长沙市望城区妇幼保健院。不幸的是，他的出生伴随着“13号染色体长臂缺失综合症”、“脑发育不良”、“虹膜缺损”等一系列缺陷和疾病。这意味着，小男孩很可能会智力障碍、生长迟缓、外表异常，几乎无法正常长大。

这个在瞬间被拖入深渊的家庭，立刻向此前负责产检的医院及华大基因展开了诘问和斥责



例：他们是否说了谎？

Accuracy 99.9%

Detection rate 70% (sensitivity)

南方周末记者发现，早在2015年12月底，华大基因便以无创产前基因检测准确率高达99.9%作为宣传点，进行全面推广。但在2018年7月13日晚间，华大基因对外发布《深圳华大基因股份有限公司关于媒体报道的澄清公告》（下称《公告》）中这一数字悄悄发生了变化。

《公告》称，“国家卫计委发布《通知》中，对于常规染色体非整倍体的检出率做了以下要求：21 三体、18 三体、13 三体的检出率分别不低于 95%、85%和 70%，而非 100%。无创基因检测作为一项筛查技术存在假阴性的风险。”



问题出在哪里？

- BGI 之前广告声称检测准确率高达 99.9%. $\rightarrow \text{Acc} = (\text{TN} + \text{TP}) / (\text{N} + \text{P}) = 99.9\%$
- 但一个不幸的家庭检测阴性却出生了13号染色体三体宝宝（trisomy 13 即 Patau syndrome）
- 之后BGI发布公告称，卫健委要求检验13号染色体三体的检出率（detection rate）不低于70%，故公司未违规 $\rightarrow \text{Sn} = \text{TP} / \text{P} = \text{TP} / (\text{TP} + \text{FN}) > 70\%$
- 讨论：BGI是否在撒谎？
- 从这件事里，检测服务提供者和用户应该吸取什么教训？



请用雨课堂讨论

Xuegong Zhang

31

投票 最多可选1项

设置

根据以上信息，你认为BGI是否在撒谎？

- ☐ A 是，之前的广告在撒谎
- ☐ B 否，一直都没撒谎
- ☐ C 是，事后的公告撒谎
- ☐ D 是，前后都在撒谎

提交

32



问题出在哪里？

- BGI 之前广告声称检测准确率高达 99.9%. ($Acc=(TN+TP)/(N+P)=99.9\%$)
- 但一个不幸的家庭检测阴性却出生了13号染色体三体宝宝 (trisomy 13 即Patau syndrome)
- 之后BGI发布公告称，卫健委要求检验13号染色体三体的检出率 (detection rate)不低于70%，故公司未违规 ($Sn=TP/P=TP/(TP+FN)>70\%$)
- Ref. according to Wikipedia, the prevalence of Patau syndrome is between 1 in 10,000 and 1 in 21,700 live births.

计算推演

- 设：待测婴儿数 $N+P=100,000$ ，其中正常 $N=99,990$ ，异常 $P=10$ ，
 - 如果 $TP=0$ ， $TN=99,990$ ， $FN=10$ ，那么 $Acc=99.99\%$ ， $Sn=0\%$ ， $Sp=100\%$
 - 如果 $TP=7$ ， $TN=99,990$ ， $FN=3$ ，那么 $Acc=99.997\%$ ， $Sn=70\%$ ， $Sp=100\%$

Xuegong Zhang

33



问题出在哪里？

- BGI 之前广告声称检测准确率高达 99.9%. ($Acc=(TN+TP)/(N+P)=99.9\%$)
- 但一个不幸的家庭检测阴性却出生了13号染色体三体宝宝 (trisomy 13 即Patau syndrome)
- 之后BGI发布公告称，卫健委要求检验13号染色体三体的检出率 (detection rate)不低于70%，故公司未违规 ($Sn=TP/P=TP/(TP+FN)>70\%$)
- Ref. according to Wikipedia, the prevalence of Patau syndrome is between 1 in 10,000 and 1 in 21,700 live births.

计算推演

- 设：待测婴儿数 $N+P=100,000$ ，其中正常 $N=99,990$ ，异常 $P=10$ ，
 - 如果 $TP=0$ ， $TN=99,990$ ， $FN=10$ ，那么 $Acc=99.99\%$ ， $Sn=0\%$ ， $Sp=100\%$
 - 如果 $TP=7$ ， $TN=99,990$ ， $FN=3$ ，那么 $Acc=99.997\%$ ， $Sn=70\%$ ， $Sp=100\%$
- 讨论：
 - BGI是否在撒谎？
 - 从这件事里，检测服务提供者和用户应该吸取什么教训？



Xuegong Zhang

34




例：他们是否说了谎？



不过，以上分析都对，但事实上这个例子里不少同一种病

万周末记者发现，早在2015年12月底，华大基因便宣称其无创产前检测准确率高达99.9%作为宣传点，进行全国推广。2018年7月13日晚间，华大基因对外发布《深圳华大基因股份有限公司关于媒体报道的澄清公告》（下称《公告》）中这一数字悄悄发生了变化。

《公告》称，“国家卫计委发布《通知》中，对于常规染色体非整倍体的检出率做了以下要求：21 三体、18 三体、13 三体的检出率分别不低于 95%、85%和 70%，而非 100%。无创基因检测作为一项筛查技术存在假阴性的风险。”



例：你是否应该去检验？

场景假设：

- 一种新发现的遗传疾病，根据群体遗传学分析，得知人群患病率为百万分之一（ 10^{-6} ），有一种检测这种疾病的方法，灵敏度是100%，特异度是99.99%，如果我去做了这个检验并且结果呈阳性，我有多大概率是真的患病？

D

怎样求解这个问题？

- 请用雨课堂弹幕回答最关键的概念/公式

Xuegong Zhang

36



例：你是否应该去检验？

场景假设：

- 一种新发现的遗传疾病，根据群体遗传学分析，得知人群患病率为百万分之一（ 10^{-6} ），有一种检测这种疾病的方法，灵敏度是100%，特异度是99.99%，如果我去做了这个检验并且结果呈阳性，我有多大概率是真的患病？

解：

- 已知 prevalence $P(D^+) = 10^{-6}$, sensitivity $Sn = 100\%$, specificity $Sp = 99.99\%$ (False positive rate $fp = 0.01\%$)
- 求 $P(D^+|T^+)$



Xuegong Zhang

37



例：你是否应该去检验？

场景假设：

- 一种新发现的遗传疾病，根据群体遗传学分析，得知人群患病率为百万分之一（ 10^{-6} ），有一种检测这种疾病的方法，灵敏度是100%，特异度是99.99%，如果我去做了这个检验并且结果呈阳性，我有多大概率是真的患病？

解：

- 已知 prevalence $P(D^+) = 10^{-6}$, sensitivity $Sn = 100\%$, specificity $Sp = 99.99\%$ (False positive rate $fp = 0.01\%$)
- 求 $P(D^+|T^+)$



$$P(D^+|T^+) = \frac{P(T^+|D^+)P(D^+)}{P(T^+)} = \frac{P(T^+|D^+)P(D^+)}{Sn \cdot P(D^+) + fp \cdot (1 - P(D^+))} \cong 1\%$$

贝叶斯公式

Xuegong Zhang

38

3.2 阈值对错误率的影响

Xuegong Zhang

39

阈值选取问题

回顾FLD中的阈值选取问题：

$$\mathbf{w}^* \propto \mathbf{S}_w^{-1}(\mathbf{m}_1 - \mathbf{m}_2)$$

- 有了投影方向，还需要确定决策的分界点

$$y = \text{sgn}\left(\sum_{i=1}^n w_i x_i + w_0\right) = \text{sgn}(\mathbf{w}^T \mathbf{x} + w_0), \quad y = \begin{cases} +1 & \Rightarrow \mathbf{x} \in \omega_1 \\ -1 & \Rightarrow \mathbf{x} \in \omega_2 \end{cases}$$

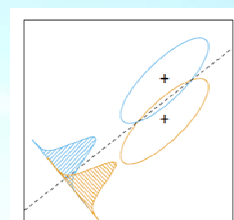
- 如何选择 w_0 ?

– 根据对数据的不同认识，可以有多种选择方法，比如

$$w_0 = -\frac{1}{2}(\tilde{m}_1 + \tilde{m}_2)$$

$$w_0 = -\tilde{m}$$

$$w_0 = -\frac{1}{2}(\tilde{m}_1 + \tilde{m}_2) + \frac{1}{N_1 + N_2 - 2} \ln \frac{P(\omega_1)}{P(\omega_2)}$$

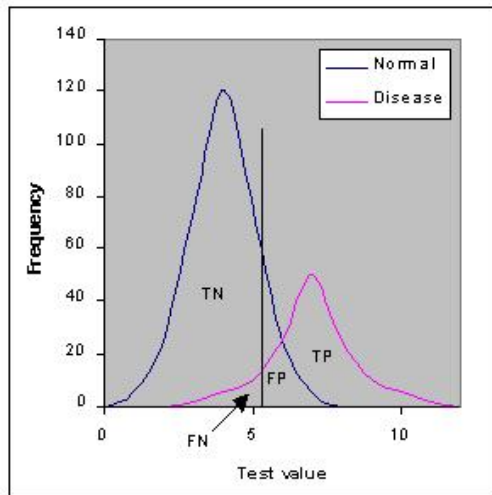
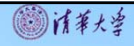


– 可以根据对错误率的要求来选择（见下周课）

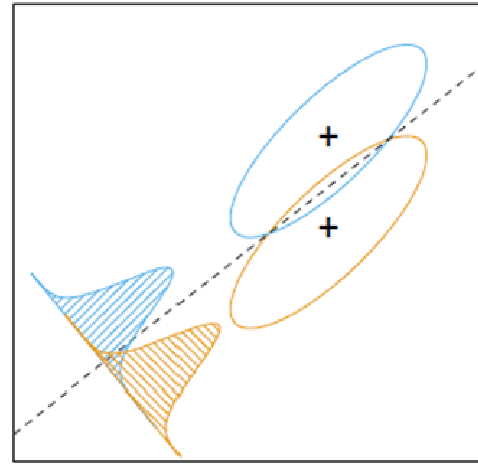
Xuegong Zhang

40

阈值选取对检测性能的影响



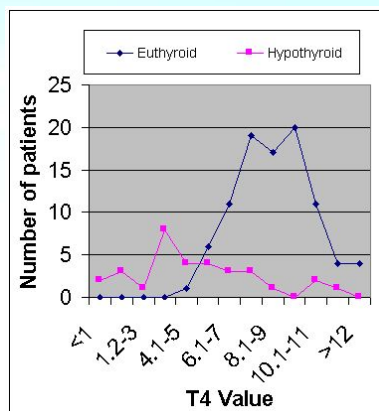
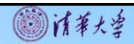
Thomas G. Tape, MD: *Interpreting Diagnostic Tests*
(<http://gim.unmc.edu/dxtests/Default.htm>), Univ. of Nebraska Medical Center



Xuegang Zhang

41

例：阈值对假阳性、真阳性的影响



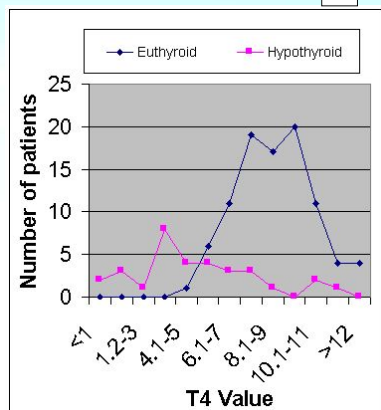
Thomas G. Tape, MD: *Interpreting Diagnostic Tests*
(<http://gim.unmc.edu/dxtests/Default.htm>), Univ. of Nebraska Medical Center

Hypothyroid: 甲状腺机能减退
Euthyroid: 甲状腺机能正常

Xuegang Zhang

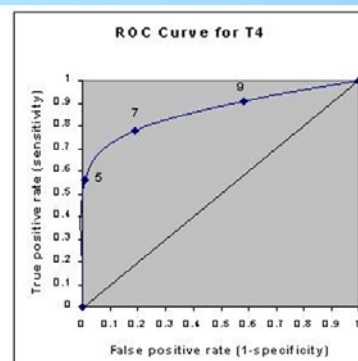
42

例：阈值对假阳性、真阳性的影响



T4 value	Hypothyroid	Euthyroid
5 or less	18	1
5.1 - 7	7	17
7.1 - 9	4	36
9 or more	3	39
Totals:	32	93

Cutpoint	Sensitivity	Specificity
5	0.56	0.99
7	0.78	0.81
9	0.91	0.42



Thomas G. Tape,
Interpreting Diagnostic
Tests, Univ. of Nebraska
Medical Center

Cutpoint	True Positives	False Positives
5	0.56	0.01
7	0.78	0.19
9	0.91	0.58

Hypothyroid: 甲状腺机能减退

Euthyroid: 甲状腺机能正常

Thomas G. Tape, MD: Interpreting Diagnostic
Tests (<http://gim.unmc.edu/dxtests/Default.htm>), Univ. of
Nebraska Medical Center

Xuegong Zhang

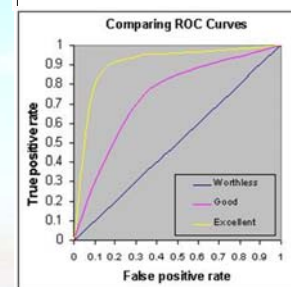
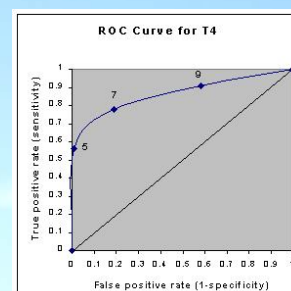
43

ROC曲线 (The Receiver Operating Characteristic curve)



The name "Receiver Operating Characteristics" came from some study in signal detection theory during World War II. Came into the medical fields in 1970's.

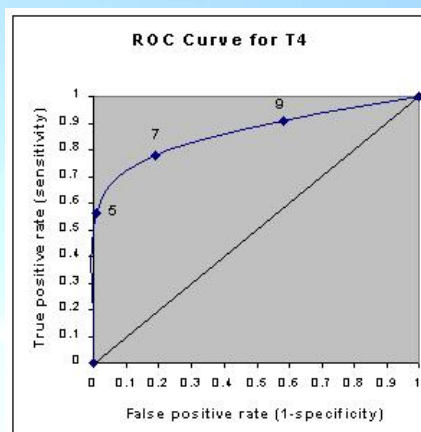
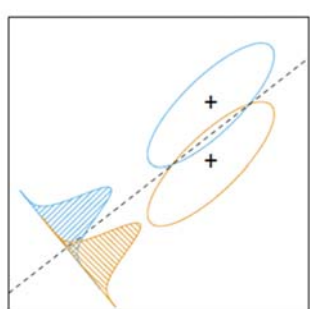
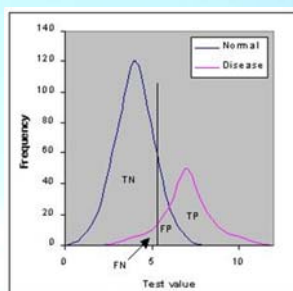
- ROC曲线反映了灵敏度与特异度之间相互制约关系
- 对于不同检测方法，曲线越靠近左上角边界则说明检测准确度越高
- 如曲线越接近左下到右上对角线，则检测越不准确
- 可以用曲线下面积（AUC）作为考虑了特异度和灵敏度均衡的检测性能指标，AUC=1是完美检测，AUC=0.5则检测完全没用
- 在评估模式识别方法性能时常常用
- 也有几种其他类似的类似形式的曲线



Xuegong Zhang

44

用ROC曲线选阈值

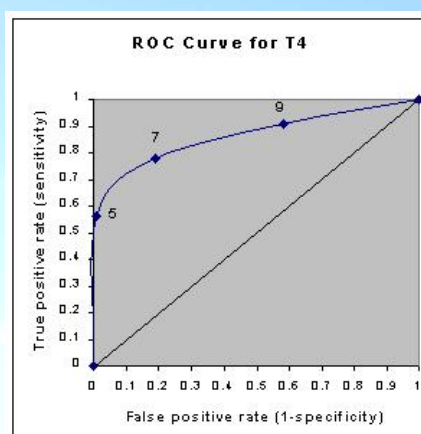
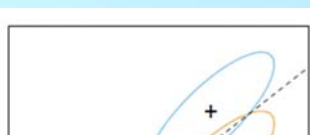
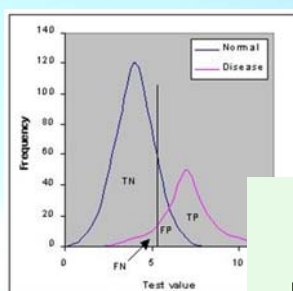


D

Xuegong Zhang

45

用ROC曲线选阈值



linear classification $h(\mathbf{x}) = \text{sign}(s)$

linear regression $h(\mathbf{x}) = s$

$s = \sum_{i=0}^d w_i x_i$

Diagram illustrating linear classification: Inputs x_0, x_1, x_2, x_d are weighted and summed to produce s , which is then passed through a sign function to output $h(\mathbf{x})$.

Diagram illustrating linear regression: Inputs x_0, x_1, x_2, x_d are weighted and summed to produce s , which is then passed through a linear function to output $h(\mathbf{x})$.

D

还有哪种学过的方法可以用ROC曲线选阈值?

请用弹幕回答

Xuegong Zhang

46



3.3 特征选择与分类器性能评估

3.4 非监督模式识别结果的评估

Let's leave it for future classes.

Xuegong Zhang

47



作业

- 第三章
 - 3.1 自学线性回归模型的评价指标
 - 3.2 新冠病毒检验的几种假想场景分析
- 截止日期：3月2日
- 第二章
 - 2.8 线性回归小实验
 - 2.9 线性分类器实验
- 截止日期：3月9日

Xuegong Zhang

48

单选题 1分

设置

休息4分钟，回到座位后请答题

- ☒ A 已回座位
- ☐ B 还没有



Xuegong Zhang



49

提交