

# 机器学习

---

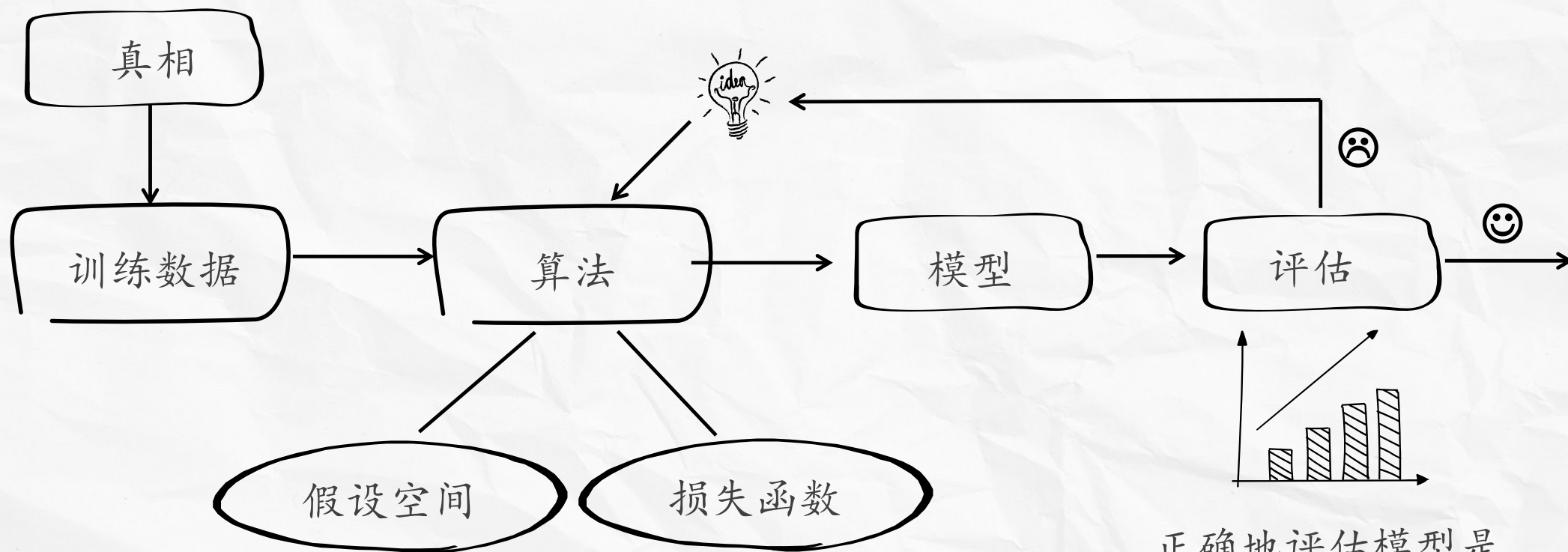
## 模型评估与选择

涂文婷

tu.wenting@mail.shufe.edu.cn

# 模型评估

## 。意义

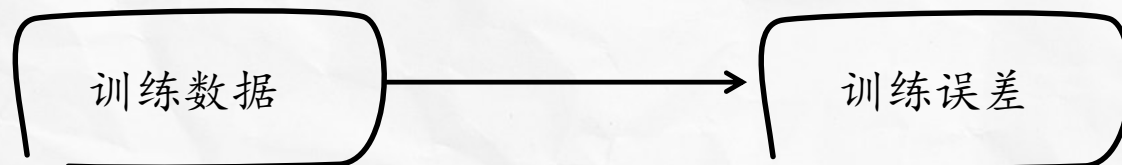


正确地评估模型是  
迭代优化模型的关键

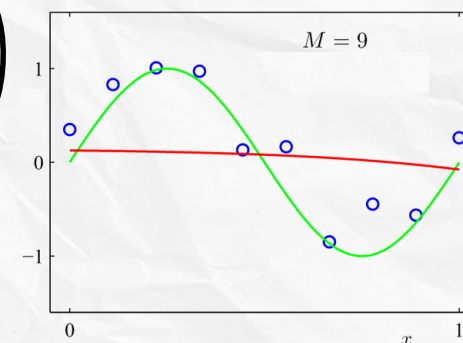
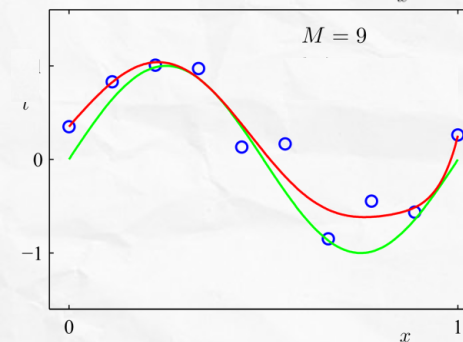
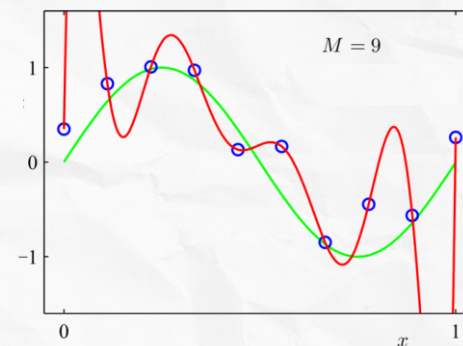
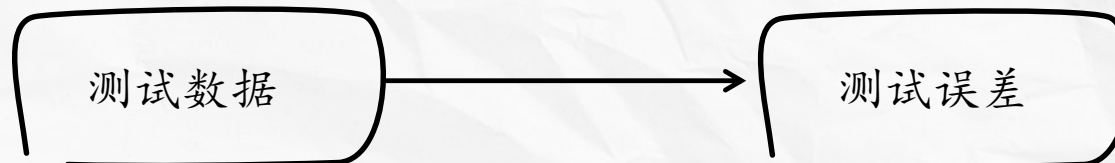
# 模型评估

## 对象

- 经验误差



- 测试误差/泛化误差

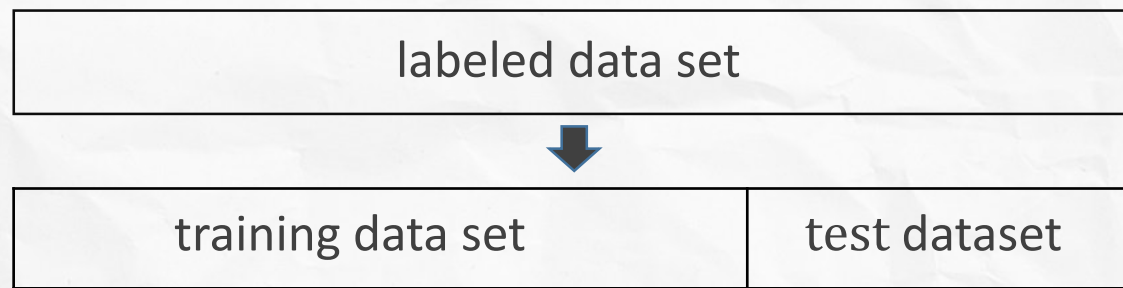


# 模型评估

---

- 训练/测试集划分

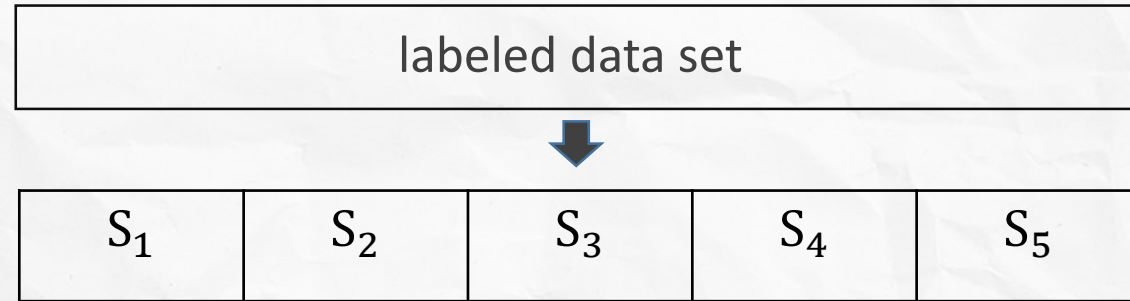
- 留出法



# 模型评估

- 训练/测试集划分

- 交叉验证法



iteration	train on	test on
1	$S_2 S_3 S_4 S_5$	$S_1$
2	$S_1 S_3 S_4 S_5$	$S_2$
3	$S_1 S_2 S_4 S_5$	$S_3$
4	$S_1 S_2 S_3 S_5$	$S_4$
5	$S_1 S_2 S_3 S_4$	$S_5$

# 模型评估

---

## ◦ 回归任务指标

- 平均绝对误差

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - f(\mathbf{x}_i)|$$

- 平均平方误差/均方误差

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - f(\mathbf{x}_i))^2$$

- 均方根误差

$$RMSE = \sqrt{MSE}$$



# 模型评估

---

## ◦ 回归任务指标

- 平均绝对误差

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - f(\mathbf{x}_i)|$$

- 平均平方误差/均方误差

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - f(\mathbf{x}_i))^2$$

- 均方根误差

$$RMSE = \sqrt{MSE}$$

# 模型评估

---

## ◦ 分类任务指标

- 错误率

$$ER = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(f(\mathbf{x}_i) \neq y_i)$$

- 正确率

$$ACC = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(f(\mathbf{x}_i) = y_i)$$



# 模型评估

## ◦ 分类任务指标

- 查准率 *Precision* 与 查全率 *Recall* (二分类场景)

真实情况	预测结果	
	正例	反例
正例	$TP$ (真正例)	$FN$ (假反例)
反例	$FP$ (假正例)	$TN$ (真反例)

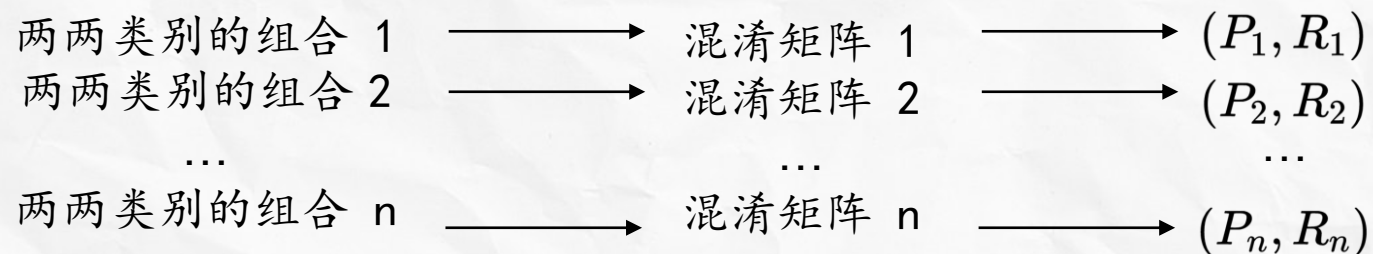
$$P = \frac{TP}{TP + FP} \quad R = \frac{TP}{TP + FN}$$
$$F1 = \frac{2 \times P \times R}{P + R} = \frac{2 \times TP}{\text{样例总数} + TP - TN}$$

# 模型评估

---

## ◦ 分类任务指标

### • 查准率 *Precision* 与 查全率 *Recall*（多分类场景）



$$\text{macro} - P = \frac{1}{n} \sum_{i=1}^n P_i$$

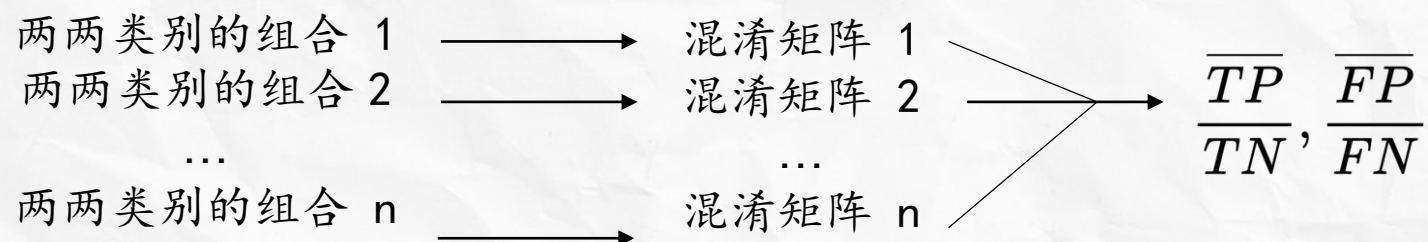
$$\text{macro} - R = \frac{1}{n} \sum_{i=1}^n R_i$$

$$\text{macro} - F1 = \frac{2 \times \text{macro} - P \times \text{macro} - R}{\text{macro} - P + \text{macro} - R}$$

# 模型评估

## ◦ 分类任务指标

- 查准率 *Precision* 与 查全率 *Recall* (多分类场景)



$$\text{micro-}P = \frac{\overline{TP}}{\overline{TP} + \overline{FP}}$$

$$\text{micro-}R = \frac{\overline{TP}}{\overline{TP} + \overline{FN}}$$

$$\text{micro-}F1 = \frac{2 \times \text{micro-}P \times \text{micro-}R}{\text{micro-}P + \text{micro-}R}$$

# 模型评估

## • 分类任务指标

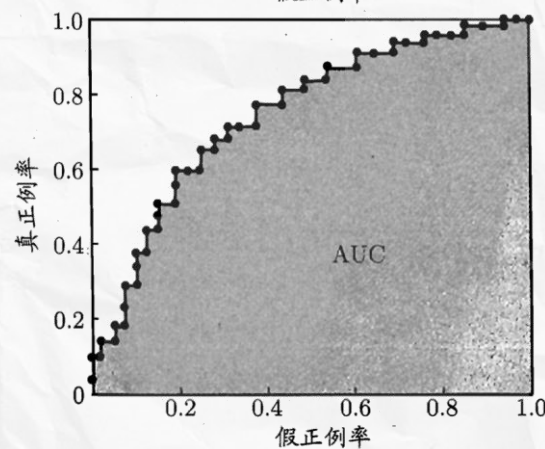
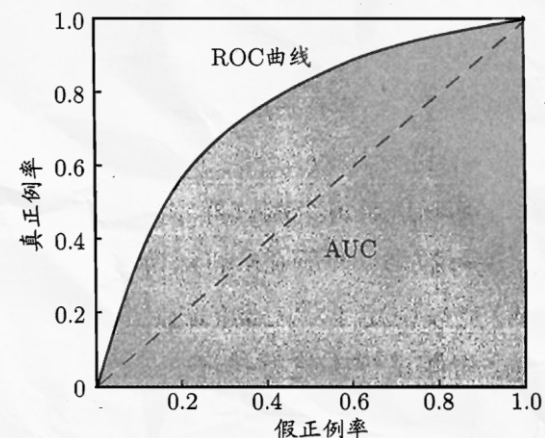
### • ROC曲线与AUC值

真实情况	预测结果	
	正例	反例
正例	$TP$ (真正例)	$FN$ (假反例)
反例	$FP$ (假正例)	$TN$ (真反例)

- “真正例率” TPR 与 “假正例率” FPR

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

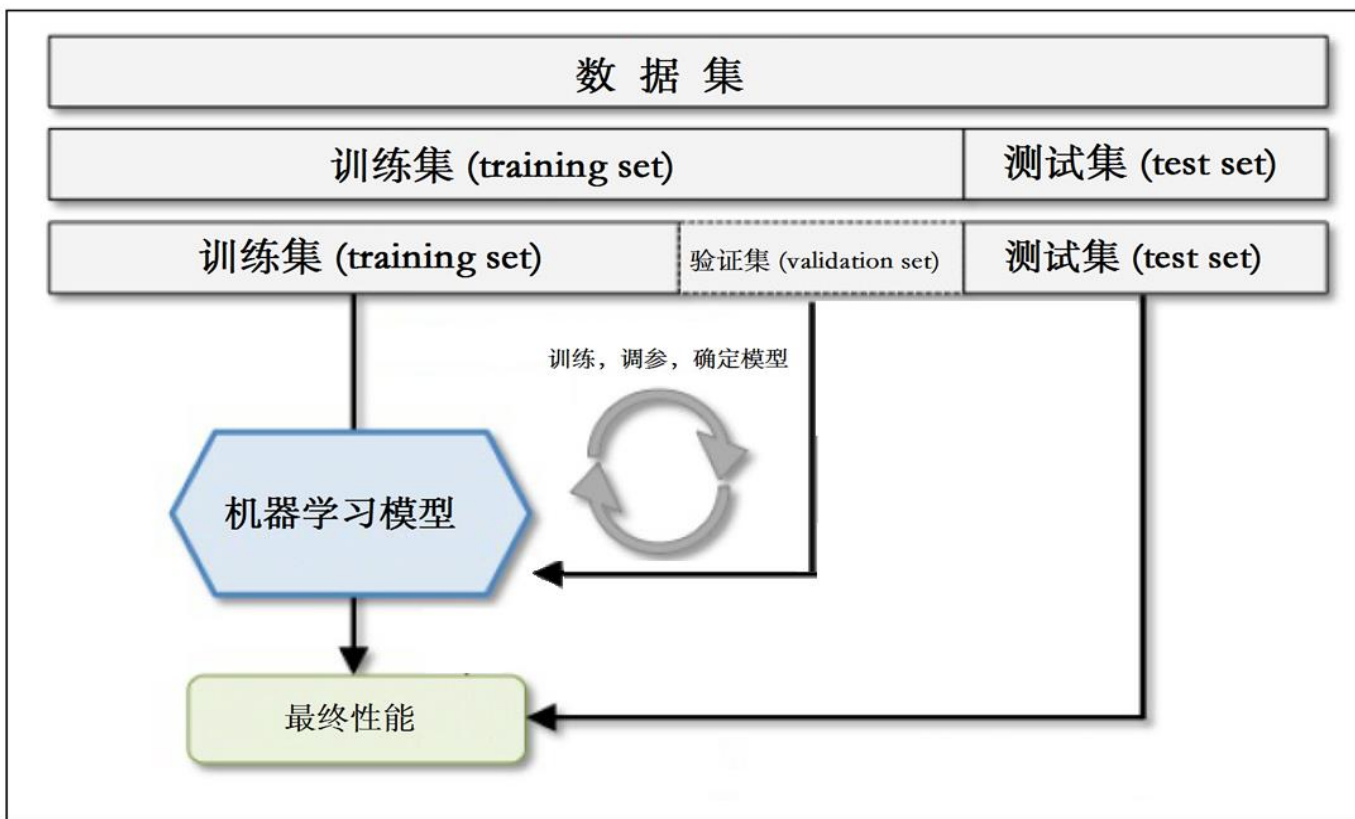
$$FPR = \frac{FP}{TN + FP}$$



AUC = “ROC 曲线之下的面积”  
= 模型给予一个随机正类样本的分数大于给予一个随机负类样本的分数的概率

# 模型选择

## “训练+验证+测试”流程



e.g., 岭回归超参数 $\lambda$

$$\min_{\mathbf{w}, b} \|\mathbf{y} - \mathbf{X}\hat{\mathbf{w}}\|_2^2 + \lambda \|\mathbf{w}\|_2^2$$