

- 2.1 采用分层抽样的方法，500个正例中含350个训练集和150个测试集，500个反例中含350个训练集和150个测试集，总共的划分方法有： $C_{500}^{150} \times C_{500}^{150}$

- 2.2略

- 2.3

BEP值：Break-Even Point 简称平衡点，满足BEP=P=R;

但学习器处于BEP值的时候，F1值可以简化为F1=P

因此若F1-A>F1-B时，有BEP-A>BEP-B

- 2.4

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

- 2.5详见「深度解析roc曲线、AUC与排序损失」一文

地址：https://blog.csdn.net/qq_41229918/article/details/126077960

- 2.6 ROC曲线的每一个点，都对应一个错误率

- 2.7 一条ROC曲线上每一点对应一个(FNR,FPR),从而确定一条代价直线，所有代价直线组合起来便构成一条代价曲线，反之亦然

- 2.8

	Min-max	z-score
优点	1、计算相对简单一点。2、当新样本进来时，只有在新样本大于原最大值或者小于原最小值时，才需要重新计算规范化之后的值	对异常值敏感低
缺点	容易受高杠杆点和离群点影响。	计算复杂

- 2.9&2.10暂时「略」，答案参考链接：<https://www.jianshu.com/p/9d70c26b73a2>】