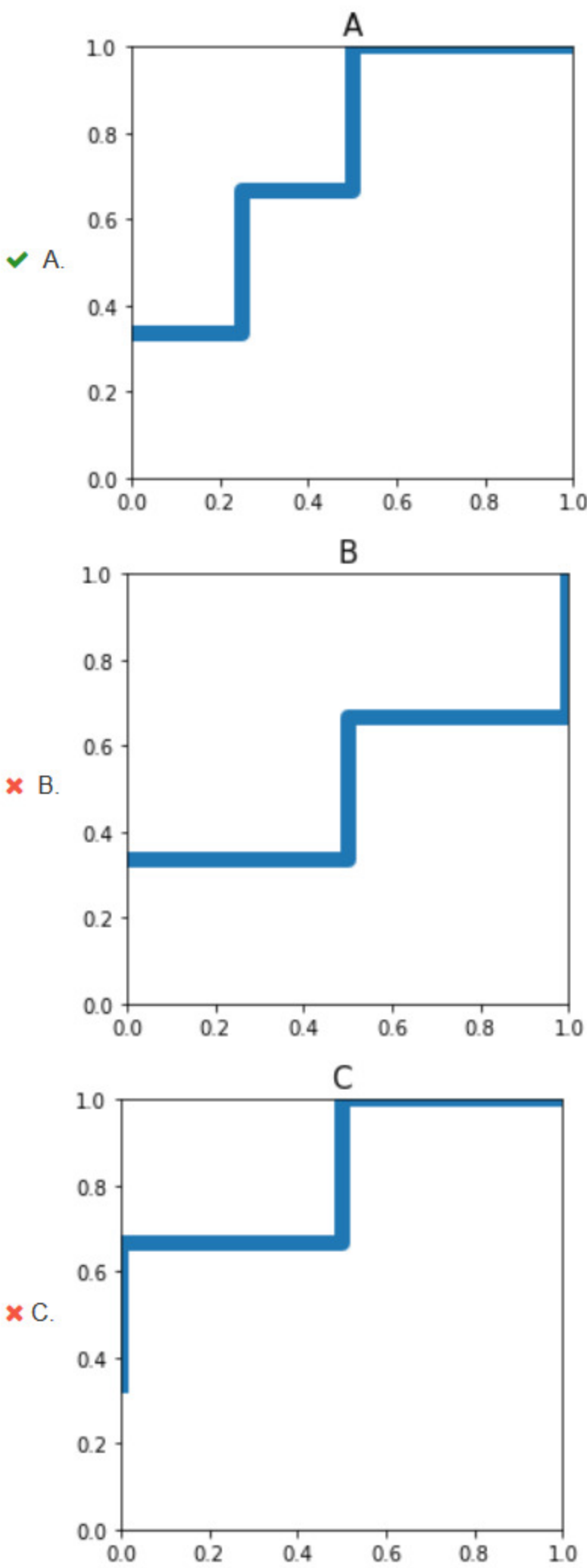


卷27：机器学习

1. 对于一组二元分类任务的测试集，其真实值为[0, 0, 0, 0, 1, 1, 1]，模型预测为1的概率为[0.3, 0.2, 0.7, 0.5, 0.4, 0.9, 0.6]，该模型在这个测试集上的ROC曲线为



解答：这道题目可以判断面试者是否真的理解ROC曲线。ROC曲线x轴为假阳性(FPR)，y轴为真阳性(TPR)。所以答案为A。

2. “近朱者赤、近墨者黑”与下面哪个算法是类似的思想？

- ✗ A. k-Means
- ✓ B. k-Nearest Neighbors
- ✗ C. Decision Tree

解答：相近的点有一样的标签，所以是k-NN模型。

3. 假如你有一个很大的训练集，而且你当前的首要考虑是训练速度，那么下面哪种正则化手段最不应该考虑？

- ✗ A.  $L_1$
- ✓ B.  $L_0$
- ✗ C.  $L_2$

解答：这三种正则项都能起到约束模型的作用。 $L_0$ 是指模型中非零参数的个数。在优化中，它没有显示的数学表达，能准确地说，含有 $L_0$ 正则项的优化问题是一个离散优化问题，而且是NP Hard。

4. 训练单个决策树是肯定无法进行并行化的。这个说法对吗？

- ✗ A. 正确
- ✓ B. 错误

解答：即使是单个决策树也是可以进行并行化的。因为每个树杈的分裂点选择上，我们需要遍历所有的特征。这个步骤是可以被并行的。

5. XGBoost与GradientBoost的不同之处不包括

- ✓ A. XGBoost使用了更复杂的决策树结构
- ✗ B. XGBoost将树的复杂度加入到了损失函数中作为惩罚项
- ✗ C. XGBoost使用了二阶导数

解答：不包括A；相反的是，XGBoost想控制（降低）树的复杂度。