

# 《数据挖掘与机器学习》期中测试题（2022-11-07）

1、已知五个点：A(0.5,1),B(2,3),C(2,0),D(2,1),E(3,2)，其中A、B为正类，C、D、E为负类，现有直线 $l_1: y+0.5x-2.5=0$ 作为初始分割线，请用感知机算法，在0.2学习率时，求该问题的正确分割直线。

2、给定一个二维空间的数据集 $T=\{(2,3),(4,7),(5,4),(7,2),(8,1),(9,6)\}$ ，已知前三个点为正类，后三个为负类，现有两点(4,4.5)和(7,4)，试用KNN算法，在 $k=3$ 和曼哈顿距离的条件下确定这两点的类型。

3、给定数据如下：

帅？	性格好？	身高？	上进？	嫁与否
帅	不好	矮	不上进	不嫁
不帅	好	矮	上进	不嫁
帅	好	矮	上进	嫁
不帅	好	高	上进	嫁
帅	不好	矮	上进	不嫁
帅	不好	矮	上进	不嫁
帅	好	高	不上进	嫁
不帅	好	中	上进	嫁
帅	好	中	上进	嫁
不帅	不好	高	上进	嫁
帅	好	矮	不上进	不嫁
帅	好	矮	不上进	不嫁

现有男生甲：帅，性格不好，身高高，不上进；男生乙：不帅，性格好，身高矮，上进，请你用朴素贝叶斯方法分别判断一下女生是嫁还是不嫁？

4、课本79页逻辑回归模型的参数估计问题，如何用524页的算法A.1（梯度下降法）求解，试给出算法的详细步骤。

5、对于第一题而言，若令直线 $l_2: y-x=0$ 和直线 $l_3: y-1.5x+1.5=0$ ，试分别计算已知的五个点关于这两条直线的函数距离和几何距离，并给出相应的支持向量和隔离带的宽度。

6、用课本100页例7.1和105页例7.2的方法求解第一题的最大间隔分离超平面和线性可分支持向量机。