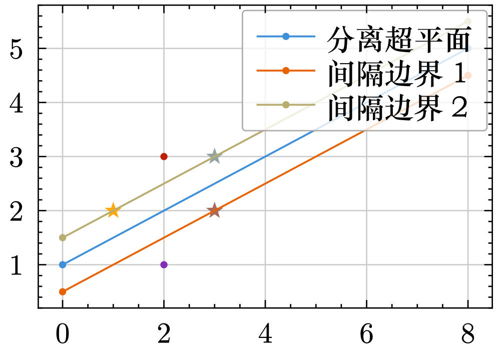
机器学习

课程作业二

redacted

1. 已知正例点 $x_1=(1,2)^T,x_2=(2,3)^T,x_3=(3,3)^T$，负例点 $x_4=(2,1)^T,x_5=(3,2)^T$，试求最大间隔分离超平面和分类决策函数，并在图中画出分离超平面、间隔边界及支持向量。
   - 支持向量为 x_1,x_3（正例）和 x_5（负例），超平面方程形式为 $w_1x^{(1)}+w_2x^{(2)}+b=0$，即 $w^Tx+b=0$，其中 $x^{(1)},x^{(2)}$ 分别为 x 的第 1 和第 2 个特征（坐标）。
   - 由支持向量可得约束条件：
     - $w_1+2w_2+b=1$
     - $3w_1+3w_2+b=1$
     - $3w_1+2w_2+b=-1$
   - 解得 $w=(-1,2)^T,b=-2$，超平面为 $-x^{(1)}+2x^{(2)}-2=0$，分类决策函数为 $f(x)=\text{sign}\left(-x^{(1)}+2x^{(2)}-2\right)$。
   - 

     - 图中三个星点为支持向量，分别为 x_1,x_3,x_5。
2. 设计单层感知机实现逻辑与、或、非运算；解释单层感知机为什么不能表示异或运算；设计多层感知机实现异或运算
   - 假设输入为 0/1，激活函数为阶跃函数 $f(x)=1 \text{ if } x\geq 0 \text{ else } 0$。
   - 逻辑与：权重 $w=(1,1)^T$，偏置 $b=-1.5$
   - 逻辑或：权重 $w=(1,1)^T$，偏置 $b=-0.5$
   - 逻辑非：权重 $w=-1$，偏置 $b=0.5$
   - 异或真值表为

输入 x_1	输入 x_2	输出 y
0	0	0
0	1	1
1	0	1
1	1	0

   - 二维平面内四个点（$\left(x_i^{(1)},x_i^{(2)}\right)$ for i in (1,4)）分布在正方形的四个顶点，其中输出为 0 的点和输出为 1 的点是两条对角线；而单层感知机对应一条直线，无法用一条直线将输出为 0 的点和输出为 1 的点分开，因此单层感知机不能表示异或运算。
   - 由于 $x_1\oplus x_2=(x_1\vee x_2)\wedge\neg(x_1\wedge x_2)$
   - 第一层计算 or 和 nand：$h_1=g(x_1+x_2-0.5),h_2=g(-x_1-x_2+1.5)$
   - 第二层计算 and：$y=g(h_1+h_2-1.5)$
3. 描述核化线性降维（如 KPCA）与流形学习之间的联系及优缺点。
   - 两者都旨在处理非线性数据的降维问题，克服了传统 PCA 只能处理线性结构的局限。某些流形学习算法（如 Isomap, LLE）可以被解释为特定核函数下的 KPCA 的特例（例如 Isomap 的核矩阵基于测地距离）

优缺点	核化线性降维	流形学习
优点	能够捕获数据的非线性结构；具有显式的投影函数，容易处理新样本；理论基础扎实（核方法）	能够处理非线性数据，能够发现数据的内在几何结构
缺点	核函数的选择（如高斯核、多项式核）没有通用标准，依赖经验；计算核矩阵复杂度高，存储开销大	对噪声和参数（如邻域大小 k）敏感；通常没有显式的映射函数，难以直接处理新样本（需要重新计算或近似）

4. 试由下表训练数据学习一个朴素贝叶斯分类器并确定 $x=(2,S)^T$ 的类标记 y。表中 $X^{(1)}, X^{(2)}$ 为特征，取值的集合分别为 $A_1=\{1,2,3\}$，$A_2=\{S,M,L\}$，Y 为类标记，$Y \in C=\{1,\ -1\}$。

特征与标记	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$X^{(1)}$	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3
$X^{(2)}$	S	M	M	S	S	S	M	M	L	L	L	M	M	L	L
Y	-1	-1	1	1	-1	-1	-1	1	1	1	1	1	1	1	-1

- 计算先验概率：
  - $P(Y=1)=\frac{3}{5}$
  - $P(Y=-1)=\frac{2}{5}$
- 计算条件概率：
  - 对于测试样本 $x=(2,S)^T$：
  - 在 $Y=1$ 类别中，$P\left(X^{(1)}=2 \mid Y=1\right)=\frac{1}{3}$；$P\left(X^{(2)}=S \mid Y=1\right)=\frac{1}{9}$
  - 在 $Y=-1$ 类别中，$P\left(X^{(1)}=2 \mid Y=-1\right)=\frac{1}{3}$；$P\left(X^{(2)}=S \mid Y=-1\right)=\frac{1}{2}$
- 计算后验概率：
  - $\mathrm{Score}(Y=1)=P(Y=1)\times P\left(X^{(1)}=2 \mid Y=1\right)\times P\left(X^{(2)}=S \mid Y=1\right)=\frac{3}{5}\times\frac{1}{3}\times\frac{1}{9}=\frac{1}{45}$
  - $\mathrm{Score}(Y=-1)=P(Y=-1)\times P\left(X^{(1)}=2 \mid Y=-1\right)\times P\left(X^{(2)}=S \mid Y=-1\right)=\frac{2}{5}\times\frac{1}{3}\times\frac{1}{2}=\frac{1}{15}$
- 由于 $\mathrm{Score}(Y=-1)>\mathrm{Score}(Y=1)$，故分类器将样本 $x=(2,S)^T$ 分类为 $y=-1$。