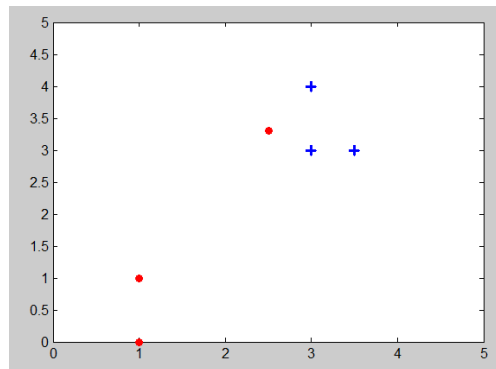


第四次作业

1. 请描述使用 EM 算法进行高斯混合模型聚类的过程。
2. 对于数据: $x_1 = (4,5)^T$, $x_2 = (1,4)^T$, $x_3 = (0,1)^T$, $x_4 = (5,0)^T$, $x_5 = (4,1)^T$, $x_6 = (0,6)^T$ 现有以下三种聚类划分:
 - (1) $\{x_1, x_2, x_6\}$, $\{x_3, x_4, x_5\}$
 - (2) $\{x_1, x_4, x_5\}$, $\{x_2, x_3, x_6\}$
 - (3) $\{x_1, x_2, x_3, x_6\}$, $\{x_4, x_5\}$假定我们聚类的准则是最小平方和误差, 请判断上述三个划分中哪个更好? 给出计算过程。
3. 请阐述 K 均值聚类 and 模糊 K 均值聚类的关系。
4. 已知正样本点 $x_1=(1,1)^T$, $x_2=(1,0)^T$, $x_3=(2.5,3.3)^T$, 负样本点 $x_4=(3,3)^T$, $x_5=(3,4)^T$, $x_6=(3.5,3)^T$, 它们的分布如下图所示



- (1) 请写出线性支持向量机需要求解的原问题和对偶问题
 - (2) 当 C 取值很大 (比如 $C \rightarrow +\infty$) 时, 定性画出会得到的决策面, 并解释原因
 - (3) 当 C 取值很小 (比如 $C \rightarrow 0$) 时, 定性画出会得到的决策面, 并解释原因
5. 结合图例, 阐述线性可分支持向量机中的支持向量的概念。