**第三次作业**

**姓名：乔翱**

**学号：20222280625**

1、考虑一个硬间隔（hard margin）支持向量机和下面来自两类的训练样本

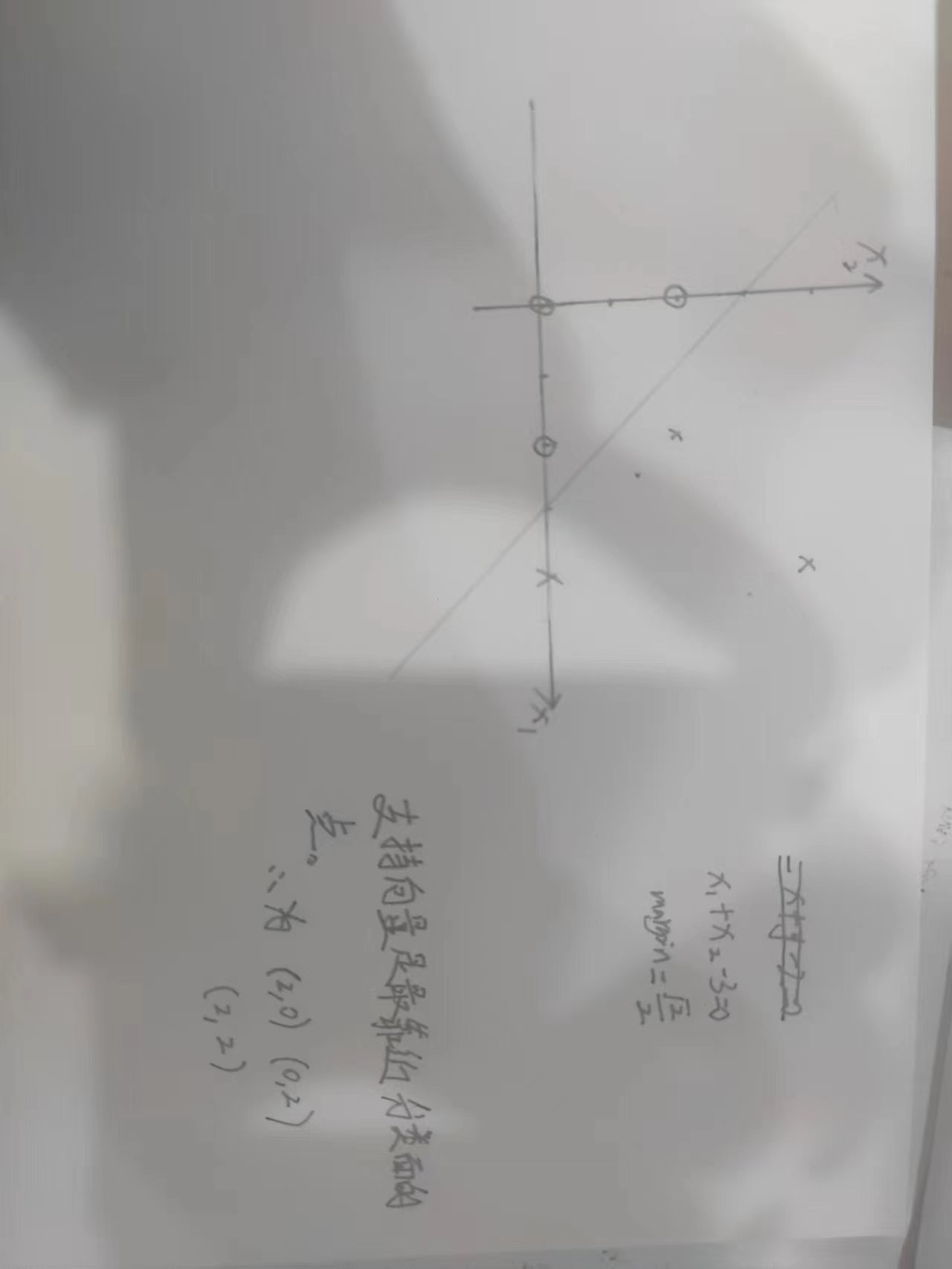
+1: (2, 2) (4, 4) (4, 0)

-1: (0, 0) (2, 0) (0, 2)

(a) 画出这6个点，通过观察画出最优分类面和权向量w，给出分类面的方程w1x1+w2x2+b=0。计算出它的间隔margin（从分类面到最近数据点的距离）。

(b) 标出所有的支持向量，并说明原因？

支持向量是训练样本中最接近分类面的样本点



2、试说明为什么核技巧（kernel trick）能让我们可以在高维特征空间运用SVM而不显著增加运行时间。

对于线性不可分的数据，通常需要把输入数据变换到高维的特征空间，使得样本线性可分。核技巧使得可以在高维特征空间中有效地学习分类边界，而不需要计算高维空间中的内积或特征向量。将内积运算的过程替换为核函数的计算，从而避免了再高维空间昂贵的计算开销。

核技巧将样本之间的内积转化为核函数的计算，核函数是一个能够直接计算样本间相似度的函数。核函数的计算复杂度仅仅与样本数相关，而与特征空间的维度无关。将样本从低维度空间映射到高维度空间，计算相似度的时间复杂度仍然可以保持与原始低维度空间相同的级别。

3、凸优化问题中的无约束优化方法梯度下降法，请列举其3种不同形式，并对比其优缺点。

批量梯度下降法，更新参数时使用所有的样本进行更新。批量梯度下降法的速度会比较，尤其是数据集非常大的情况下，收敛速度就会非常慢，但是由于每次的下降方向为总样本的平均梯度，得到的会是一个全局最优解。

随机梯度下降法。每一步仅仅选取一个样本来求梯度。训练速度很快，即使在样本量很大的情况下，可能只需要其中一部分样本就能迭代到最优解，由于每次迭代并不是都向着全体样本的最优化方向，导致梯度下降的波动非常大，更容易从一个局部最优跳到另一个局部最优，，准确度下降。

小批量梯度下降法，每一步采用一部分固定数量的样本来计算。小批量梯度下降法既保证了训练的速度，又能保证最后的收敛的准确率。

当训练集比较小时，批量学习，采用牛顿或者共轭梯度下降。当训练集大时，使用随机梯度下降。当训练集基于两者之间，使用小批量梯度下降。

4、简述凸优化问题中的等式约束优化及不等式约束优化方法。

对于凸优化问题，需要找最小化一个凸函数的变量，在等式约束中，还需要满足一个或多个等式约束；而在不等式约束中，需要满足一个或多个不等式约束。

等式约束，约束条件会将解的范围限定在一个可行域，此时不一定能找到使得为**∇xf(x)**为0的点，只需要找到在可行域内使得f(x)最小的值即可。对于带有等式约束的优化问题，可以使用拉格朗日乘数法将等式约束转化为拉格朗日乘数形式，从而得到一个无约束的优化问题。

当约束区域包含目标函数原有的可行解时，此时加上约束可行解仍落在约束区域内部对应g(x)<0的情况，这时约束条件不起作用。对于带有不等式约束的优化问题，可以使用投影梯度法、内点法或外点法等方法求解。将约束问题转换为无约束问题进行求解。