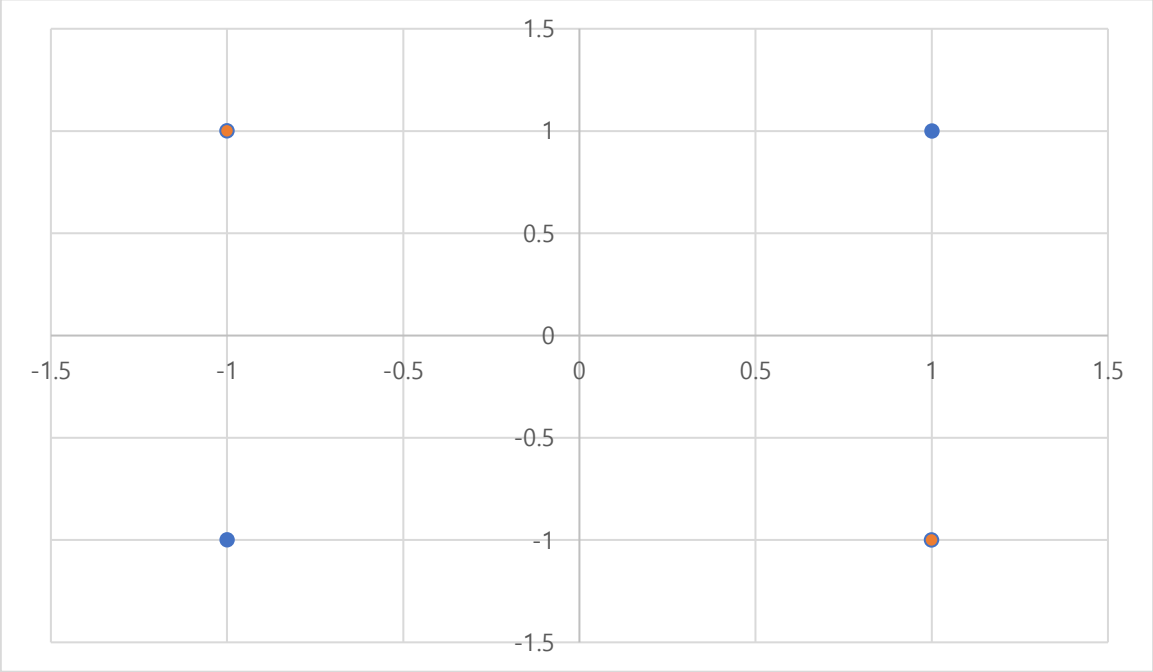
2 Theory Questions

2.1 SVM

(a)

No. This is a typical XOR problem.



In the graph above, blue dots represent X_+, orange dots represent X_-. We can clearly see that we cannot draw "a" line to separate the positive examples from negative examples.

(b)

X_+ :

$\phi\big((1,1)\big) = (1,1,1,1)^T$

$\phi\big((-1,-1)\big) = (1,-1,-1,1)^T$

X_- :

$\phi\big((1,-1)\big) = (1,1,-1,-1)^T$

$\phi\big((-1,1)\big) = (1,-1,1,-1)^T$

(c)

Put $w = (a,b,c,d)^T$. Then in expression

$$\min ||w|| \quad s.t\ \ y_i(w^T x_i) \geq 1\ for\ all\ i$$

$$L = \frac{1}{2}w^T w - \sum_{i=1}^{4} \lambda_i (y_i w^T \phi(x_i) - 1)$$

$$\frac{\partial L}{\partial w} = 0 \;\Rightarrow\; w = \sum_{i=1}^{4} \lambda_i y_i \phi(x_i) \;\Rightarrow\; w^T w = \sum_{i=1}^{4} \lambda_i y_i w^T \phi(x_i) = \sum_{i=1}^{4}\sum_{j=1}^{4} \lambda_i \lambda_j y_i y_j \phi(x_i)^T \phi(x_i)$$

$$Q(\lambda) = \sum_{i=1}^{4} \lambda_i - \frac{1}{2}\sum_{i=1}^{4}\sum_{j=1}^{4} \lambda_i \lambda_j y_i y_j \phi(x_i)^T \phi(x_i)$$

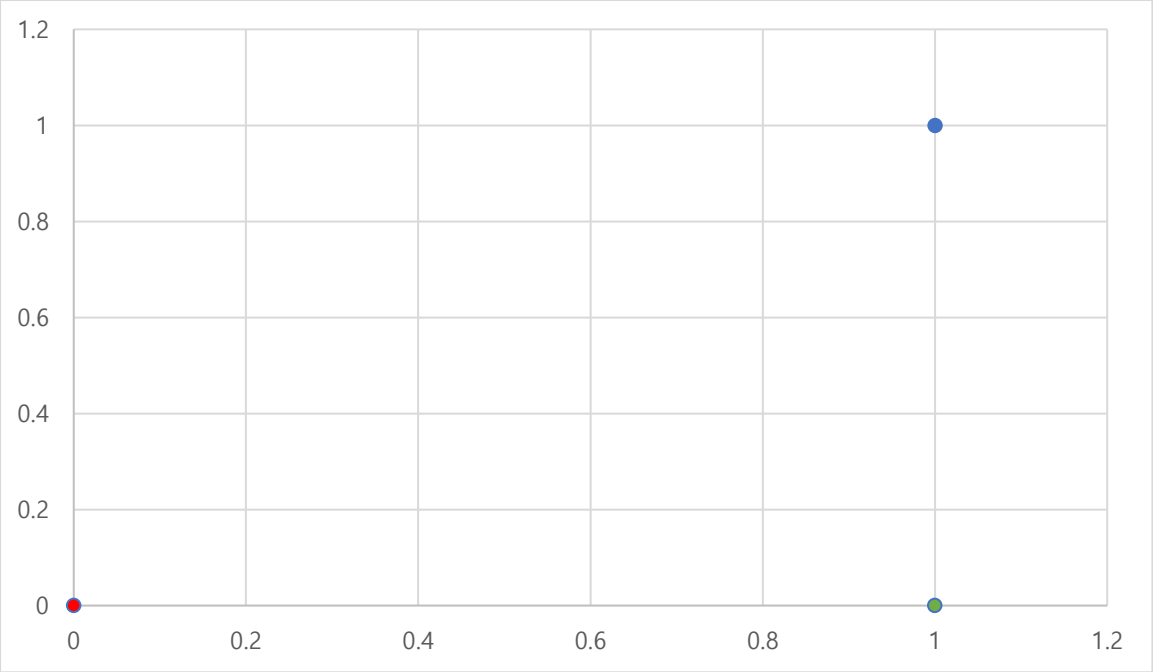
$$Q(\lambda) = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 - \frac{1}{2}(\lambda_1\lambda_1 * 4 + \lambda_2\lambda_2 * 4 + \lambda_3\lambda_3 * 4 + \lambda_4\lambda_4 * 4)$$

$$Q(\lambda) = \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 - 2{\lambda_1}^2 - 2{\lambda_2}^2 - 2{\lambda_3}^2 - 2{\lambda_4}^2$$

$$\therefore \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \frac{1}{4}$$

$$w = \sum_{i=1}^{4} \lambda_i y_i \phi(x_i) = \frac{1}{4}(1,1,1,1)^T + \frac{1}{4}(1,-1,-1,1)^T - \frac{1}{4}(1,1,-1,-1)^T - \frac{1}{4}(1,-1,1,-1)^T = (0,0,0,1)^T$$

2.2 K-means



Suppose k=2, and we have total $2\gamma + 1$ data, γ data on (0, 0), γ data on (1,0), and one datum on (1, 1). The optimal clustering example would be when the cluster centers are at (0, 0) and (1, 0). In this case the cost of optimal 2-means clustering would be $1^2 = 1$.

However, assume that in greedy initialization method, in step 1 we picked (1, 1) as the first cluster center. Then in step 2 (0, 0) will be picked as the second cluster center. Then there will be γ data left on (1, 0), thus the cost of 2-means clustering would be γ^2, which is at least γ times larger than the cost of the optimal 2-means clustering.

3 Programming questions

5.

If you use SIFT, the number of detected visual words for images will change a lot. For example, if one image is very uniform, then there will be very small number of interest points detected. This is because SIFT extracts interest points such as corners or edges. In contrast, denseSIFT uses grid to extract visual words, therefore, there will be little variance of numbers of detected visual words compared to SIFT.