1. (10 points) John has the following dataset:

Class	x1	x2
A	0	0
A	1	0
A	0	1
B	1	1
B	2	2
B	2	0

He is studying machine learning and encountered some questions which he is unable to solve. Help him solve the following questions.

1. (2 points) Are the points linearly separable? Support your answer by plotting the points.
2. (3 points) Find out the weight vector corresponding to the maximum margin hyperplane. Also find the support vectors present.
3. (2 points) What is the effect on the optimal margin if we remove any one of the support vectors in this question?
4. (3 points) In general, for any dataset, what can we say about the effect on optimal margin, if we remove any of the support vectors?

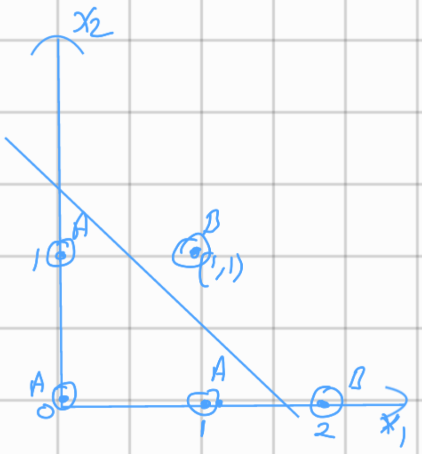
Q1) Ans) According to question

→ Hyperplane 1: for class A: $\frac{x_2 - 1}{x_1 - 0} = \frac{0-1}{1-0}$

$x_2 - 1 = -x_1 \Rightarrow \boxed{x_1 + x_2 = 1}$

→ Hyperplane 2:

$\frac{x_2 - 1}{x_1 - 1} = \frac{2-1}{0-1} \Rightarrow \frac{x_2 - 1}{x_1 - 1} = -1 \Rightarrow \boxed{x_1 + x_2 = 2}$



Equation of decision boundary: $x_1 + x_2 = \frac{3}{2}$

Q2) Ans) Maximum Margin Hyperplane

→ we know distance between hyperplane to closest sample of either class is maximum

Class A: $[(1,0),(0,1)] \rightarrow x_1 + x_2 = 1$

Class B: $[(1,1),(2,0)] \rightarrow x_1 + x_2 = 2$

Distance b/w lines: $\frac{|C_2 - C_1|}{\sqrt{a_1^2 + b_1^2}} = \frac{|2-1|}{\sqrt{2}} = \frac{1}{\sqrt{2}}$

→ we know that $L_1 \parallel^{el} L_2$ and $L_1 \parallel^{el} H$ and $L_2 \parallel^{el} H$

[here 'H' = decision boundary]

Now $\frac{2}{\|w\|} = \frac{1}{\sqrt{2}} \Rightarrow \|w\| = 2\sqrt{2}$ (1)

Let a vector on w be (K,K) so $\cancel{\sqrt{2}}K = 2\cancel{\sqrt{2}}$

$K = 2$ (2)

so $w = [2,2]$ and $b = -\frac{3}{2}K \Rightarrow b = -\frac{3}{2} \times 2$

then $\boxed{b = -3}$

We can say that equation of hyperplane: $2x_1 + 2x_2 = 3$

$\Rightarrow \boxed{x_1 + x_2 = 3/2}$

Q3) According to question

→ if any one of the 4 support vector is removed

1) There will be no change in optimal margin

→ if both of A → [(1,0) & (0,1)] or both of B [(2,0) & (1,1)] are removed, then the optimal margin increase

→ if all the 4 support vector are removed, the optimal margin will increase

Q4) a) According to question

→ if we remove any of the support vector in general

*) when we drop some support vector in maximization problem we will get an optimal value because the set of support vector satisfying the original [stronger set of contraints is a subset of the candidate satisfying the new set of constraints]

For the weaker constraints, there may be additional sol^n that are even better. In other words

$$\boxed{\max_{x \in A, x \in B} f(x) \leq \max_{x \in A} f(x)}$$

In SVM, we maximize the margin with respect to the given trainning points. where we drop constraints. it can increase or stay the same [depend on the data set]

In general, it is expected that the margin increases, when we drop the support vectors.