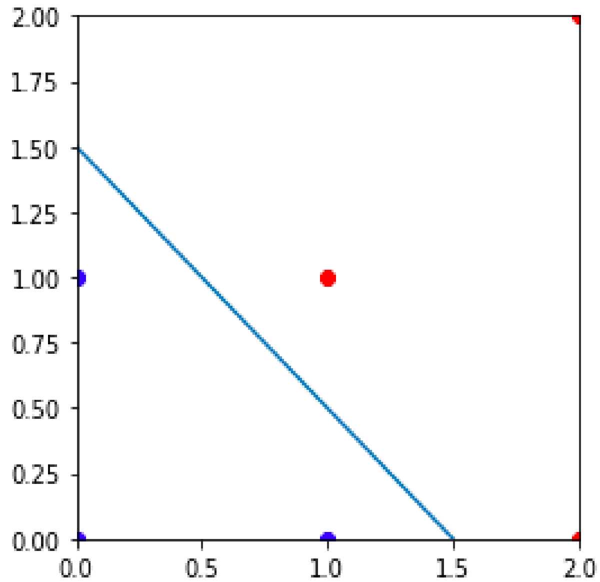
# Homework 2

**1** Consider a binary classification problem. Let us denote the input set as $x_n$,$y_n$ where $x_n$ is the data point and $y_n \in \{1, \text{-}1\}$ is the corresponding class label. Further, misclassifying a positive class data point costs k times more than misclassifying a negative class data point. Rewrite the SVM optimization function to model this constraint.

**Answer**.Minimize $1/2w^T w + kC\sum_{i=0}^{n} E_p + C\sum_{i=0}^{n} E_n$ where $E_p$ and $E_n$ are errors of positive and negative classification points (we penalize the positive classification points $k$ times more)

**2** **a.** Plot these six training points. Are the classes $\{+, -\}$ linearly separable?

**Answer.**Yes, they are linearly separable. We can check by inspection, examining this graph below.




**b.** Construct the weight vector of the maximum margin hyper-plane by inspection and identify the support vectors.

**Answer.** The 4 support vectors are **[1, 0]**, **[0, 1]**, **[1, 1]**, **[2, 0]**. The weight vector of the maximum margin hyper-plane is [-3, 2, 2] as [x0, x1, x2] where x0 is the threshold and the hyper-plane is given by $x_1 + x_2 = 1.5$. The margin is given by 1/norm(w) = 1/($2^2$+$2^2$) = 0.3536.

**c.** If you remove one of the support vectors, does the size of the optimal margin decrease, stay the same, or increase? Explain.

**Answer.** If we remove [1,0] or [1,1] the size of the optimal margin increases as the weight vector becomes [-3,2,0] with the margin increasing to 0.5 from 0.3536. However, if we remove any of the other points, by inspection, the hyperplane remains the same.

**d.** Is your answer to **c.** also true for any dataset? Provide a counter-example or give a short proof.

**Answer.** No. For this dataset, if we remove [0,1], the optimal margin stays the same/increases. However, for large datasets, removing one support vector shouldn't affect the margin much especially if the support vector is clustered close to other support vectors, as the relevant constraints will still be there.

**3.** A kernel is valid if it corresponds to a scalar product in some (perhaps infinite dimensional) feature space. Remember a necessary and sufficient condition for a function $K(x, x)$ to be a valid kernel is that associated Gram matrix, whose elements are given by $K(x_n, x_n)$, should be positive semi-definite for all possible choices of the set x. Show whether the following are also kernels:

**a.**$K(\mathbf{x}, \mathbf{x}') = c\langle \mathbf{x}, \mathbf{x}' \rangle$

**Answer.** Yes, if c is positive, then this is a valid kernel. We can rewrite this kernel as $\langle \sqrt{c}\mathbf{x}, \sqrt{c}\mathbf{x}' \rangle = c\langle \mathbf{x}, \mathbf{x}' \rangle$; not possible when c is lesser than 0.

**b.**$K(\mathbf{x}, \mathbf{x}') = \langle \mathbf{x}, \mathbf{x}' \rangle^2 + \mathbf{e}^{-\|\mathbf{x}\|^2}\mathbf{e}^{-\|\mathbf{x}'\|^2}$

**Answer.** Yes.$(\mathbf{x}, \mathbf{x}') = \langle \mathbf{x}, \mathbf{x}' \rangle$ is a valid kernel as seen from the previous part (taking c=1), so its square is also a valid kernel. Moreover, $\mathbf{e}^{-\|\mathbf{x}\|^2}\mathbf{e}^{-\|\mathbf{x}'\|^2}$ is always positive (exponent raised to the negative power of norm of $x$ is always positive), so the sum is also a valid kernel.