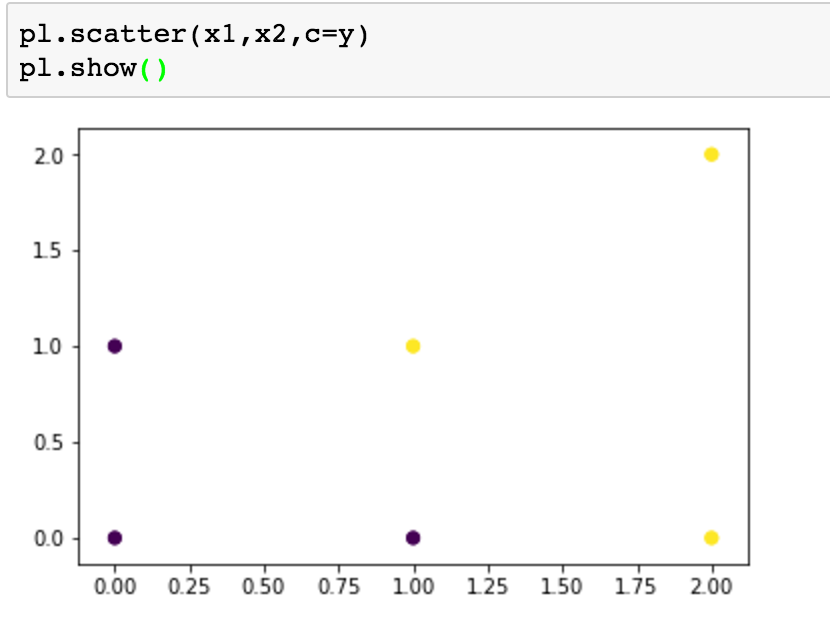
CS3244 Homework2 Essay Questions

1. Consider a binary classification problem. Let us denote the input set as {***x****n*,*yn*} where ***x****n* is the data point and *yn* ∈ {-1,+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.

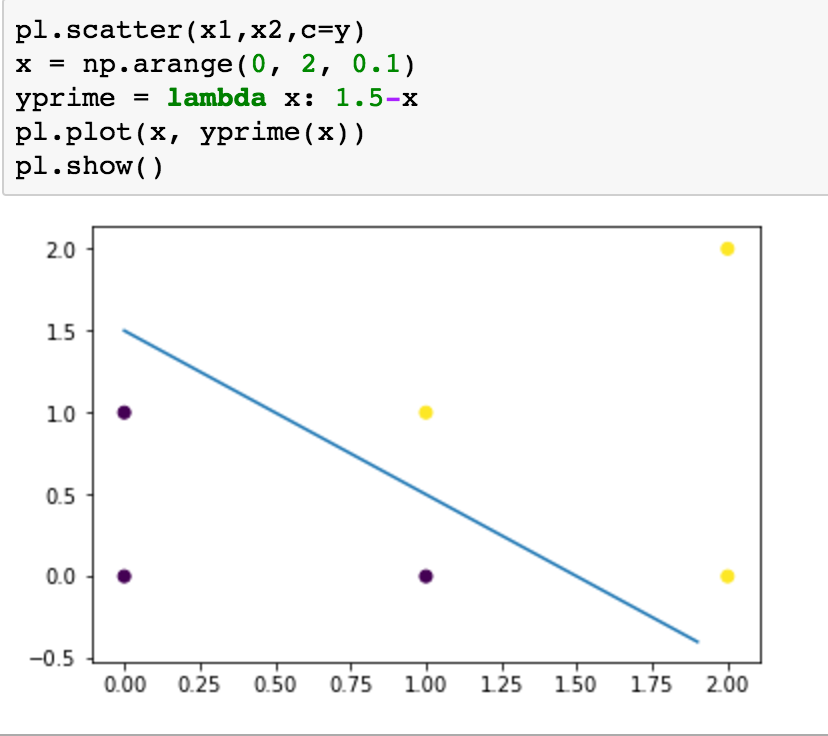
Answer:

1. Consider the following training data:
   1. Plot these six training points. Are the classes {+, −} linearly separable?

|  |  |  |
| --- | --- | --- |
| X1 | X2 | Y |
| 1 | 1 | 1 |
| 2 | 2 | 1 |
| 2 | 0 | 1 |
| 0 | 1 | -1 |
| 1 | 0 | -1 |
| 0 | 0 | -1 |



Yes, they are linearly separable.

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

The maximum margin hyper-plane’s weight is given by: (1, 1)

The bias b is given by -1.5

Support vectors are given by these four (x1, x2):

(1, 0), (0, 1), (1, 1), (2, 0)

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

It depends.

If you remove (1, 0) or (1, 1), the size will increase from 1/√2 to 0.5

If you remove (0, 1) or (2, 0), the size of the optimal margin will stay the same.

* 1. Is your answer to (c) also true for any dataset? Provide a counterexample or give a short proof.

True, the size of the optimal margin will be increasing or staying the same with less number of support vectors given (some of them (removed).

Proof:

Assume we have a support vector, if removed, the optimal margin decrease. Then it means adding this vector will increase the margin. By the SVM algorithm, the constraint is equivalent to None as it doesn’t decrease the margin, which means we can safely ignore it, and it is not a support vector, producing contradiction.

1. 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(xn,xm)*, should be positive semi-definite for all possible choices of the set **x**. Show whether the following are also kernels:
   1. *K(****x****,****x****') = c <****x****,****x****'>*

It is a kernel.

Consider the Gram matrix corresponding to this kernel, then the determinant of the matrix is always 0, making the matrix itself always positive semidefinite.

* 1. *K(****x****,****x****') = <****x****,****x****'>2 +****exp(-x^2)exp(-x’^2)***

It is a kernel.

Consider the two-dimensional scenario i.e x = {x1,x2} , the determinant of the gram matrix for this can be derived as (x2\*exp(-x1^2)+x1\*exp(-x2^2))^2, which is nonnegative. This proves that the gram matrix is always positive semidefinite.