1. The leave-one-out error for the first instance is 0. Since, first instance is not a support vector, and therefore does not affect the optimal hyper-plane if left out.

For instances 2 and 3, we firstly calculate slack variables, and. We also know that the maximum length of feature vector is 1, that is,. We can then calculate the necessary conditions of leave one out error as:;. Thus, instance 2 cannot produce a leave-one-out error, while instance 3 could produce a leave-one-out error.

Thus, the upper bound of leave-one-out error for the three instances is 1.

1. From the dual training algorithm, in each iteration of the algorithm, we will add 1 to if. From the problem statement, we know that any two feature vectors are orthogonal, then we know that the summation,, could be simply expressed as . That is, we are adding 1 to only when. We notice that when starts at 0, the inequality will be satisfied, and if, the inequality will never be satisfied. Therefore we will add 1 to, and never change the values in the later iterations. Thus, the corresponding to is 1, for all.
2. We claim that C should be changed by a factor of so the new solution would define the same linear classifier. We denote the scaled version with \*, then we claim:

We can see this from the dual problem. In the dual problem, we defined the objective function as:

We could multiply a factor of to the objective function, and the optimal solution of should not change. That is, if we change the objective function to the following, the optimal solution should be exactly the same as :

Now consider the modified version of the problem, where we set , the new dual problem can be represented as:

Subject to:

This could be expressed as:

Subject to:

We immediately notice that from (\*), is a solution to this maximization problem with optimal value . We also notice that since we scaled the bound of decision variables by of the original problem, by linearity, one could not get a solution better than . Thus, this indeed generates the optimal solution to the new dual problem.

Then, we could conclude that . Therefore, if we change to , the optimal hyperplane is still the same as before. That is, it is still the same classifier. The resulting