# Optimization

## Assignment 2

Yonatan Simson, 015764921

Amit Kristal, 025602624

### Q1

In the above figure we wish find d, the minimal distance between point and line .

We do so by projecting onto , using the principal of orthogonality:

QED

### Q2

What makes this problem convex is the minimization of L2 norm in

In fact this is equivalent to the quadratic programming problem as learned in class:

Prove: is convex

Proof:

Due to triangle inequality of the L2 Norm

QED

Prove: defines a convex set

Proof:

### Q3

Condition (4a) is constraint given in the Primal form.

To prove the rest we use KKT conditions. The Lagrangian for this problem is:

Differentiating (\*) by and will give us

We of course have from the Lagrangian

Proof of (4c)

Say we have the following Lagrangian of a general convex problem ()

Then the first order optimality condition is

If the constraint is inactive then

If the constraint is active then

Either way the equality holds true.

Using this we get:

The active constraints correspond to the case of

Where as the inactive constraints correspond to the case of

So we conclude that only the vectors corresponding to the active constrains play a role in defining . These are *the* support vectors that give SVM it's name.

### Q4

The Dual form is derived by taking the results from the previous question (4d), (4e) and substituting them into the Lagrangian:

Using strong duality we derive the dual form from the above equation:

QED

### Q5

If we have then we know that exists. If it does exist then we know that where is the collection of that fulfil the conditions of problem 4. This means that strong duality holds according Slater's condition. Being that if the feasible region has a viable point then:

Equation (4d) can be expressed in matrix form as:

So extracting ***w*** from is all we need to do.

In order to extract we make use of the inequality constraints when they are active.

We can also average over the entire active set to get a better estimate

### Q6

For the following problem

s.t.

Condition (4a) is constraint given in the Primal form.

To prove the rest we use KKT conditions. The Lagrangian for this problem is:

The Lagrange multipliers:

The KKT conditions:

And the existing inequality constraints:

Differentiating (\*\*) by , and will give us:

We shall look at three different cases of :

Case 1: for when points are correctly classified,

The above points will not have any effect on the solution

Case 2: for when points are correctly classified but within the margin

Case 3: for when points are on the wrong side of decision boundary

The points in both cases have an effect on the penalty and will make up the set of support vectors

Coupled with the condition of Lagrange multiplier we get the following box constraint:

### Q7

Substituting the above results will give us the dual Lagrangian:

Therefore the dual problem is given as:

### Q8

As before the weights are:

Finding requires some more work. It is defined solely by the support vectors when . This leads to, using the complementary slackness:

If this implies that . Using we get that . These are points exactly on the margin line

If this implies that . This means that matches point correctly labeled but within the margin and matches the points incorrectly classified.

For support vectors obeying the following condition we know that

By averaging over the entire set:

Same as in separable case.

### Q9

The cost function with a hinge loss is convex but not differentiable at every point:

The above is the PEGASOS variant of the cost function

The subgradient for this function is:

Define the set of incorrect and within the margin samples:

So each update involves iterating through:

Where

See Matlab appendix for the calculation of the subgradient

### Q10

For the following box set

The projection of x on

### Q11

Matlab

### Q12

For the following problem:

The Lagrangian is

Deriving the KKT optimality conditions

If both are zero then this means both constraints are non active. This means that:

If only and then is an active constraint:

If only and then is an active constraint:

### Q13

For the following problem

I will approach this problem in this way. If I knew the solution I would know which of the constraints are active the problem would be easy to solve. But I do know what the solution to the unconstrained problem would be. . Using this can expose which constraints would be active . The since it is inactive.

Now we can use the Lagrangian to solve this problem analytically.

The final answer is:

When starting at the gradient at this point is:

The Hessian is: (Unitary) since we can write our function as

The since this is the required step size to get to from using

According to this . This just so happens to be the correct solution for our problem. The reason for this would be that the Hessian is Unitary matrix.

In order to make this point we can take a different example. Such as:

Where

The correct solution is found by using the projected Newton algorithm as described in the question 11. If however we use without using the reduced Hessian we get:

Instead of the true constrained optimal solution:

### Q14

The equation for the augmented Lagrangian for problem (10) is:

Where

This leads to:

Remainder:

|  |
| --- |
| **Algorithm: Augmented Lagrangian**   * Init: , * Iterate:   + Minimize , using the projected Newton   + Update : , |

### Q15

The stopping condition I used was:

The tolerance I used was . The regularization factor was C=1

The rate of convergence for the Projected Newton was slower than the Gradient Projection. It took 81sec instead of 64sec for Gradient Projection. I hypothesis this is due to the box constraints being a more significant factor than the optimal step direction

### Q16

The accuracy rate for Train set is 100% using both methods. The accuracy for the test set using Newton Projection is: 99.1906%, For Gradient Projection: 99.0894%