Intro To ML – HW3

students:

Dor Bank - 301380416 - [dorbank@gmail.com](mailto:dorbank@gmail.com),

Avi Caciularu - 203056585 - avi.c33@gmail.com

**Theory Questions**

**Question 1:**

We will show a lower bound: . let be a unit vector from the standard basis of size , where '' appears in the place.

We will begin by showing that the set can be shattered by . In order to show that, we need to show that every label assignment for is consistent with some hypothesis in . let be the assignment for the vectors . Then define hypothesis as:

.

Furthermore, for the computation of the result at the output node we will set:

We get

Note that here we assume that . the permutation for is trivial.

**Question 2:**

In this questions we will use the following lemmas, proved in recitation 7:

Lemma 1:

Lemma 2:

Lemma 3:

1. Denote the function family of each node *j* in layer *i* as . In our case, since the node separators are independent of each other, we get that for each layer *i*, . Because there are d nodes at each layer, we can see that all are equal to our wanted . From Radon's theorem, we get that for each , . from Lemma 2 we get that
2. is simply composed from concatenating the family times. Explicitly:

. From Lemma 3 we get that

1. For every node, we have parameters for the weight vector w, and more for the bias . In total, we have parameters, layers with nodes, and layer (the final one) with one node. Combining all together we get

.

1. Assume . First notice that . otherwise we get which is surely not true for a large enough .

Now, by taking the log of each size we get:

, where the second transition is true as stated above.

1. first notice that . Now, let there be a set of size that is shattered. Then . combining this with the inequality mentioned we get . from section d we get that . Thus, we conclude that .

**Question 3:**

1. Consider the new primal problem of SVM:

The regular SGD step was:

Denote so

We will show that:

And then the new SGD with the “Subgradient Projection Method” step would be:

The definition of the projection of on ( ) is .

If then the trivial solution is , so we will assume that :

* If so and then .
* If so and we get that the normalized vector and then , because is the optimal solution for .

1. Let be a convex set, and .

Because is a convex set, there is so that .

Let . Particularly, this equality is correct for : Hence . From the Pythagoras’s theorem, we get that meaning that . Hence, we get by the triangle inequality.

1. We have the same requirements of the original proof. the minimizer demands since it is in the last iteration. Hence, the expression in the equation after (12) fulfills: , because we minimize the norm of the vector by applying the ball projection on it, so all the expression’s norm should be smaller.

After that we continue the analyze and get:

and after that (only after equation (13)), the rest of the proof remains the same, and Theorem 1.1 still holds for stochastic Subgradient Projection Method.