Tohar Lukov 304956824

Natan Elul 305402398

Ass2

1.

a)

We want to minimize the problem:

$$\lambda\|w\|^2 + \sum_{i=1}^{m}\left[\ell^h\big(w,(x_i,y_i)\big)\right]^2$$

The equivalent problem:

$$minimize\ \lambda\|w\|^2 + \sum_{i=1}^{m}\xi_i^2$$
$$s.t. \forall i, y_i\langle w, x_i\rangle \ge 1-\xi_i\ and\ \ \xi_i \ge 0$$

$\ell^h\big(w,(x_i,y_i)\big) = \max\{0, 1-y_i\langle w,x_i\rangle\}$

- If $y_i\langle w,x_i\rangle \ge 1$ by the minimization process -> $\xi_i = 0 = \ell^h$
- If $y_i\langle w,x_i\rangle < 1$, by the minimization process -> $\xi_i = \ell^h$ the minimization including the quadratic value.

b)

$objective\ z\ will\ be: (w, \xi_1, \xi_2, \ldots\ \xi_m)^T$
$Id_{\dim m} = identity\ matrix\ in\ size\ \dim m \times \dim m$

$$H = 2\begin{pmatrix} \lambda \cdot Id_d & 0 \\ 0 & Id_{\dim m}\end{pmatrix}$$

$$u = (0, \ldots, 0)^T \qquad \big(size(u) = (1, m+d)\big)$$

Denote $y_i \cdot x_i$ as the vector $\big(y_i \cdot x_i(1),\ y_i \cdot x_i(2), \ldots, y_i \cdot x_i(d)\big)$

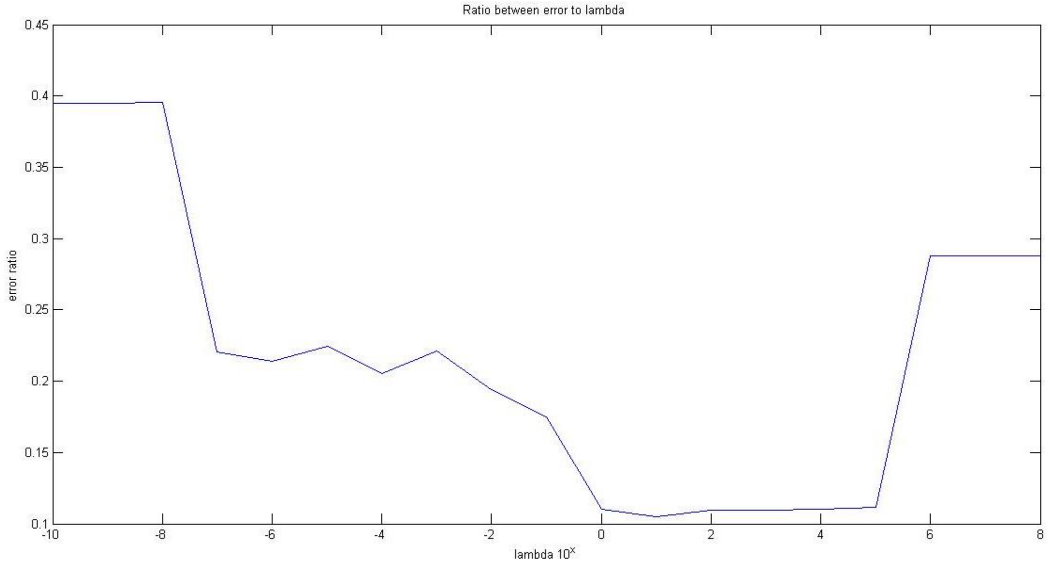
$$A = \begin{pmatrix} y_1 \cdot x_1 & 1 & 0 & 0 & 0 \\ y_2 \cdot x_2 & 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_m \cdot x_m & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$size(A) = (2m, m + d)$$

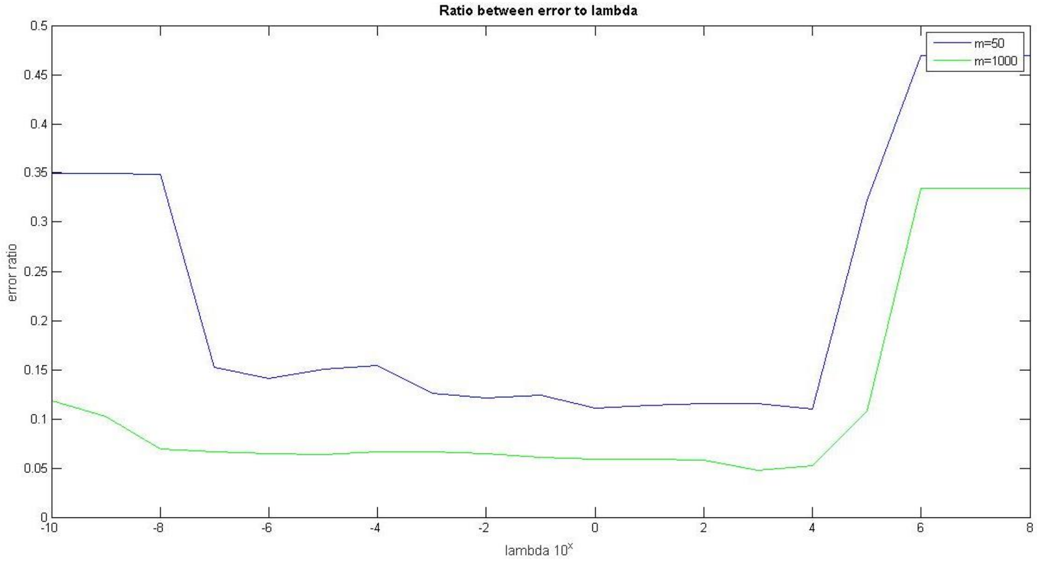
$$v = (1_1, 1_2, \dots, 1_m, 0_{m+1}, 0_{m+2}, \dots, 0_{2m})^T$$

3.

a)



b)



c)

We saw in the class that the λ value is the tradeoff between the margin size and minimizing the hinge lost error component.

The error equation of soft svm from the class:

$$\mathbb{E}_{S\sim\mathcal{D}^m}\left[err\left(h_{\widehat{w}_S},\mathcal{D}\right)\right] \leq \min \ell^H(u,\mathcal{D}) + \lambda\|u\|^2 + \frac{2R_{\mathcal{D}}^2}{\lambda m}$$

Small λ:

For $m = 50$ we can see from the graph that for small λ values the error is quite big, and we associate it to the margin size (small margin for high λ), because a small margin size can lead to false detect future samples (noise on the samples, noise on the current measure…).
But for a bigger sample size (m = 1000), even if we have small margin, the error is still low.

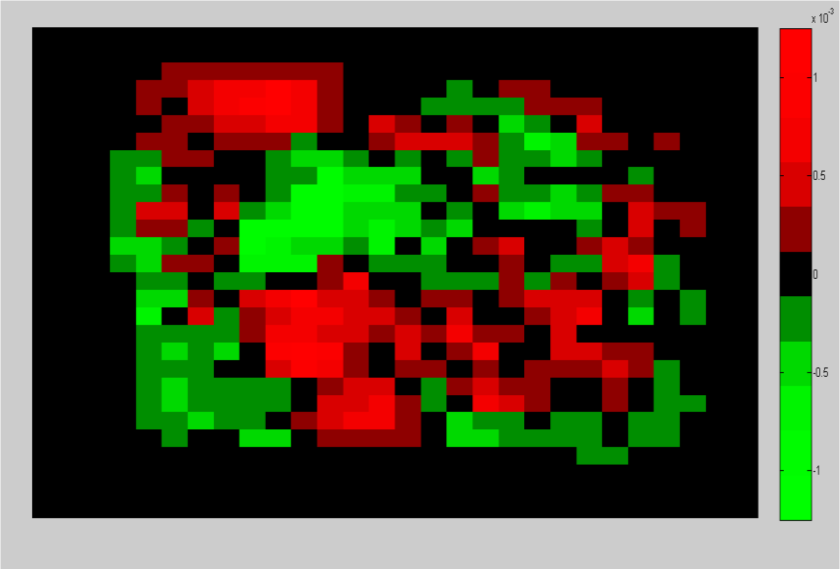
We can look at the equation and understand why it's true $\left(\frac{2R_{\mathcal{D}}^2}{\lambda m}\right)$.
As we saw in the class, we can attribute this error to "*estimation error*", as long the margin is small and we need a lot more samples to overcome that error, and we see how big sample size is helping.

Big λ:
For both $m = 50$ and m=1000, we can see that the error increased, and we can
We can also see that for a bigger sample size (green line) the error is lower, and again, by looking at the equation there is division by m.

We can explain this phenomenon by "*approximation error*", as long we are decreasing the size of H (Large margin).

The similarity between the lines are that both of them have bigger error on the sides (very large margin, and very small margin), and it logics according the error formula.

d)



e)

The 3 shape examples tagged as -1 label, while 5 shapes tagged as 1.

This shape demonstrates that the values of the w coordinates are corresponding to the values of the examples coordinates. Green pixels (means negative w coordinates) represents the significant coordinates in the 3 shape, while the red pixels (means positive w coordinates) represents the significant coordinates in the 5 shape.

We can see that the significant coordinate in the shapes, presents as a darker color.

The explanation is that w was built in a purpose that the dot product of it and the example x would be positive if x represented a five shape, and negative if it represented a three shape.

Recall that each example consist of coordinates with the value of 1 or 0, depending on the type of the shape it represents.

Hence, the coordinates of the 5 shape will multiplied with the corresponding positive value coordinates of w, results a “plus” tagging.

In the same way, the coordinates of shape 3 will multiplied by more negative coordinates of *w* (the green coordinates), and the summation of them will properly be negative and classified as 3.

In addition, we can see that the shared coordinates of the 3 shape and the 5 shape, and the coordinate within the border, has zero value (black colored), as those coordinates don’t have an effect on the classification and are more neutrals.

4.

a)

We will prove it by induction.

Base case:

$$t = 0$$

We need to show that:

$$|w^1(i)| \leq t$$

By perceptron algorithm:

$$w^{(1)} = (0, 0, 0 \dots, 0)$$

and for each coordinate i it is hold that:

$$|w^1(i)| = 0 \leq t$$

Assumption:

Assume that for some iteration $k = t$ the inequality holds:

$$\left|w^{(t+1)}(i)\right| \leq t$$

Step:

We need to show inequality hold for:

$$\left|w^{(t+2)}(i)\right| \leq t + 1$$

Observation:

a. For each coordinate on vector x_j it is hold that $\left|x_j\right| \leq 1$, and for each index j it is hold that $\left|y_j\right| = 1$.
b. From (a), for each i it is hold that $\left|y_j x_j(i)\right| \leq 1$.

Proof:

Let's take a look on the $w^{(t+2)}$ vector, that created by the perceptron algorithm.

On each iteration update the value of $w^{(t+2)}$ changes as follow:

$$\boldsymbol{w^{(t+2)}} = \boldsymbol{w^{(t+1)}} + \boldsymbol{y_j x_j}$$

Induction hypothesis:

$$\left|w^{(t+1)}\ (i)\right| \leq t.$$

Observation:

$$\left|x_j y_j(i)\right| \leq 1$$

And by the Triangle inequality:

$$\left|w^{(t+2)}(i)\right| = \left|w^{(t+1)}\ (i) + \left|y_j x_j(i)\right|\right| \leq \left|w^{(t+1)}\ (i)\right| + \left|y_j x_j(i)\right| \leq t+1$$

b) We want to prove that $|w^T(i)| \geq 2^{i-1}$ for each $i \leq d$.

We will show it by induction that for all $i \leq d$ exists that: $w^T(j) \geq 2^{j-1}$

Base:

We need to show that it is valid for the coordinate $i = 1$:

As we showed before, because w separate **all** the samples in S, by separating x_1 it exists that:

$$w^T\ (1) \geq w^0\ (1) + \ y_1 x_1(1) = 0\ +\ 1 = 2^{1-1} = 2^0$$

Hypothesis:

Let assume that for each coordinate $j\ \ s.t.\ \ \ j < i \leq d,$ it is hold that

$$w^T(j) \geq 2^{j-1}$$

(without the absolute value. It's following that w is positive)

Proof:

Show on each value of i:

We need to separate to cases:

Even i:

Let's take a look on vector x_i, because i is even, the vector x_i looks like:

$$[1_1, 1_2, \dots, 1_{i-1}, -1_i, 0_{i+1}, \dots 0_{\boldsymbol{d}}]$$

By the samples S definition, and because w labels correctly all the examples in S, it is hold that:

$$y_i \langle w, x_i \rangle \geq 1$$

Because i is even, $y_i = -1$.

$$-1 \cdot \left(\sum_{j=1}^{i} w^T(i) \cdot x(i) \right) \geq 1$$

$$-1 \cdot \left(w^T(i) \cdot x(i) + \sum_{j=1}^{i-1} w^T(j) \cdot x(j) \right) \geq 1$$

$$-1 \cdot \left(w^T(i) \cdot (-1) + \sum_{j=1}^{i-1} w^T(j) \cdot 1 \right) \ge 1$$

$$w^T(i) - \sum_{j=1}^{i-1} w^T(j) \ge 1$$

$$w^T(i) \ge \sum_{j=1}^{i-1} w^T(j) + 1$$

$$w^T(i) \ge 1 + \sum_{j=1}^{i-1} w^T(j) \ge^1 \; 1 + 2^0 + 2^1 + \dots + 2^{j-1} \; = 2^j - 1 + 1 = 2^{i-1}$$

1– by the induction hypothesis

We prove that for any even i it exists that $w^T(i) \ge 2^{i-1}$

<u>*Odd i:*</u>

Because i is odd, the vector x_i look like:

$$[-1_1, -1_2, \dots, -1_{i-1}, \qquad 1_i, 0_{i+1}, \dots 0_{\boldsymbol{d}}]$$

By the samples S definition, it is hold that

$$y_i \langle w, x_i \rangle \ge 1$$

Because i is odd, $y_i = \; 1$.

$$\left(w^T(i) \cdot x(i) + \sum_{j=1}^{i-1} w^T(j) \cdot x(j) \right) \ge 1$$

$$w^T(i) - \sum_{j=1}^{i-1} w^T(j) \ge 1$$

$$w^T(i) \ge \sum_{j=1}^{i-1} w^T(j) + 1$$

And we got the same equation as in the even case.

We show that $w^T(i) \geq 2^{i-1}$ and by the properties of an absolute value:

$$|w^T(i)| \geq w^T(i) \geq 2^{i-1}$$

c) from (a), in any iteration, and for any $i \leq d$, in our sample:
$|w^{t+1}(i)| \ \leq t$
And from (b), for every coordinate I, $|w^T(i)| \geq 2^{i-1}$.

$$T - 1 \geq |w^T(i)| \geq 2^{i-1}$$

And for the max coordinate $i = d$

$$T - 1 \geq |w^T(i = d)| \geq 2^{d-1}$$

$$T \geq 2^{d-1} + 1$$

Therefore, the number of iteration is exponential by d.
But in the real life, the algorithm is way worse:$2^9 \ll 349526$

d	Number of iterations
1	1
2	6
3	22
4	86
5	342
6	1366
7	5462
8	21846
9	87382
10	349526