AI501 Homework 2

20200130 Yujun Kim

March 2023

1 Exercise #2.3

(a) According to figure1, two calsses are linealy seperable. Black dots are for positively labeled data, and yellow dots are for negatively labeled data. For example, $x = 1.5$ can give linear boundary.

(b) I perform some new arguments for finding support vectors of the SVM. This may need some care to understand:

In (c), we obtained a *feasible solution* of the primal problem, $w = (0, 2), b = 1$, which gives the primal function value $p = 2$. In (d), we obtained a *feasible solution* of the dual problem, $a_1 = 1, a_2 = 1, a_3 = 0, a_4 = 2, a_5 = 0, a_6 = 0$, which gives the dual function value $d = 2$. By weak duality, this implies that $p* = d* = 2$, where $p*, d*$ are optimal values of primal and dual problem respectively. Thus, give w, b and $a\prime_s$ are solution of primal and dual problem respectively. In particular, $I^+ = \{1, 2\}, I^-\{4\}$(refer (2.23), (2.24) of textbook). i.e. support vectors are x_1, x_2, x_4

(c) I use symbols in the textbook. From the definition of support vectors,

$$< w, (2, 1) > +b - 1 = 0 \tag{1}$$

$$< w, (-2, 1) > +b - 1 = 0 \tag{2}$$

$$< w, (1, 0) > +b + 1 = 0 \tag{3}$$

Figure 1: Problem 2.3

(1) - (2) gives $< w, (4,0) >= 0$. Thus, $w = (0, k)$ for some k. Insert into (2), (3) and obtain

$$k + b - 1 = 0, b + 1 = 0$$

Thus, $b = 1, k = 2$ and $w^* = (0, 2)$ so that $p^* = \frac{1}{2}||w^*||^2 = 2$.

(d) $a_3 = a_5 = a_6 = 0$. The dual formulation is

$$\max\ f = a_1 + a_2 + a_4 - (3a_1a_2 - 2a_1a_4 - 2a_2a_4) - \frac{1}{2}(5a_1^2 + 5a_2^2 + a_4^2) \quad (4)$$

$$\text{with } a_1 + a_2 - a_4 = 0, a_i \geq 0 \quad (5)$$

By Lagrange multiplier method, unless the solution is on boundary, the optimal solution satisfies

$$\begin{bmatrix} \partial f/\partial a_1 \\ \partial f/\partial a_2 \\ \partial f/\partial a_4 \end{bmatrix} = k \begin{bmatrix} \partial g/\partial a_1 \\ \partial g/\partial a_2 \\ \partial g/\partial a_4 \end{bmatrix}, \text{ for the constraint } g = a_1 + a_2 - a_4$$

This leads to

$$1 - 5a_1 - 3a_2 + 2a_4 = k \quad (6)$$

$$1 - 5a_2 - 3a_1 + 2a_4 = k \quad (7)$$

$$1 - a_4 + 2a_1 + 2a_2 = -k \quad (8)$$

which we conclude to $a_1 = a_2 = 1, a_4 = 2, k = -3$. Let $a = 1, a_1 = a_2 = a, a_4 = 2a$. Then, by (4),

$$d^* = f^* = 4a + 5a^2 - 7a^2 = 4a - 2a^2 = 2$$

2 Exercise #2.4

(a) According to figure2, two classes are not linearly separable. Black dots are positively labeled and yellow dots are negatively labeled.

(b) Every boundary was of the form $x = k$, and k for various C(0.2*n for $n = 1, \cdots 10$ is given in the figure3. k grows by C. i.e. the decision boundary moves right as C increase.(see figure 4)

(c) From figure3, we conclude the boundary converges to $x = 0.75$ as C goes to infinity.

3 Exercise #2.5

(a) According to figure5, two classes are not linearly separable. Black dots are positively labeled and yellow dots are negatively labeled.

(b) Let $\varphi(x) = (||x||^2, 0, 0) \subseteq \mathbb{R}^3$. Then,

$$\varphi(x_i) = \begin{cases} 18 & \text{for } i = 1, 2, 3 \\ 2 & \text{for } i = 4, 5, 6 \end{cases}$$

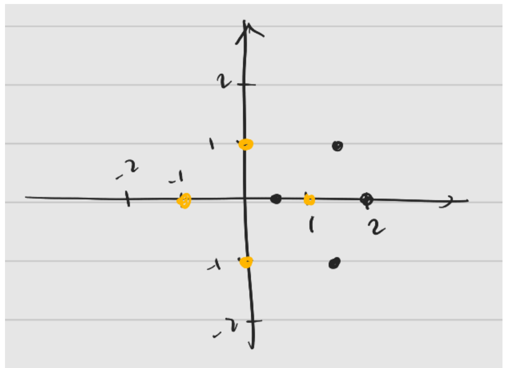

Figure 2: Problem 2.4

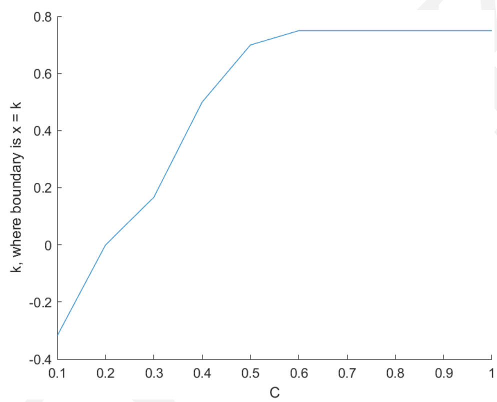

Figure 3: k according to C

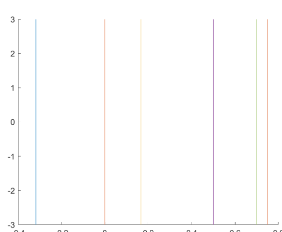

Figure 4: Different decision boundaries

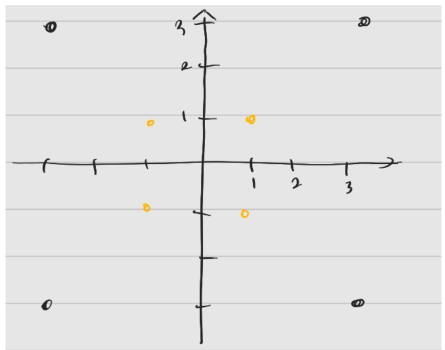

Figure 5: Problem 2.5

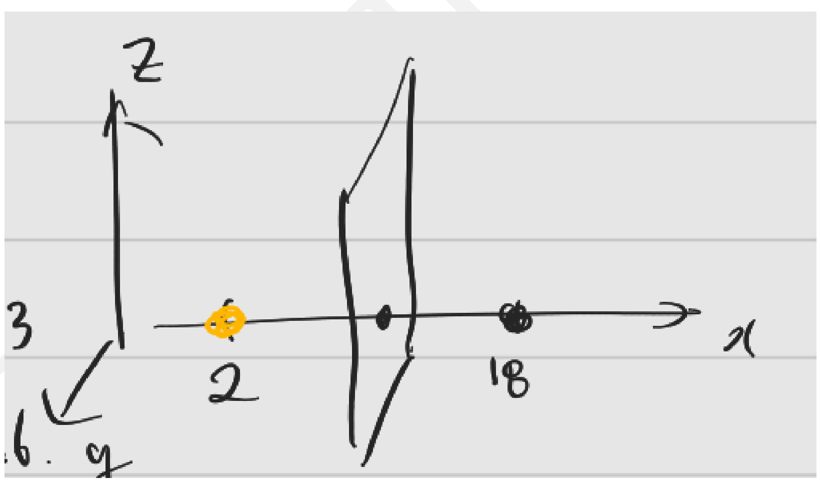

Figure 6: Feature space

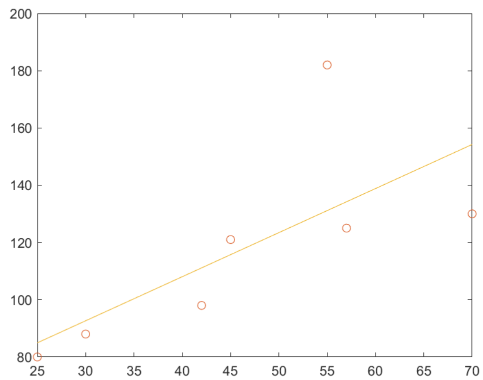

Figure 7: OLS for blood pressure per age

Hence, $x = 10$ in the feature space separates two classes. This is shown in figure 6. The colors are preserved by the transform φ

(c) The kernel function is $k(x, y) = \varphi(x)\varphi(y) = ||x||^2||y||^2$

(d) $\varphi(x_1) = \varphi(x_2) = \varphi(x_3) = (18, 0, 0), \varphi(x_4) = \varphi(x_5) = \varphi(x_6) = (2, 0, 0)$. For maximizing margin between two data points with different label, the hyperplane that bisects the line segment connecting two data points is the boundary. Thus, $(18, 0, 0), (2, 0, 0)$ both have to serve as support vector in the feature space.

Let $a = a_1 + a_2 + a_3, b = a_4 + a_5 + a_6$, where a_i are variables in dual formulation. Note that $y_i y_j < \varphi(x_i)\varphi(x_j) >$ only depends on the labels of x_i, x_j. Thus, the dual problem of SVM, (2.16) in the textboook, reduces to

$\max_{a,b} a + b - \frac{1}{2}(a^2 \times 18^2 + b^2 \times 2^2 - 2 \times 2 \times 18 \times ab$ subject to $a, b \geq 0, a - b = 0$

From $a = b$, we have

$$\max_a 2a - \frac{1}{2}(324a^2 + 4a^2 - 72a^2) = -128a^2 + 2a, a \geq 0$$

which FONC gives $-256a + 2 = 0$. Thus, $a = b = 1/128$. Then, $d^* = -128a^2 + 2a = 1/128$ is the solution of the dual problem.

4 Exercise #3.2

(a) By performing OLS estimate using MATLAB, we obtain formula

$$y(mmHg) = 46.4414 + x(age) \times 1.5398$$

(b) Figure 7 shows OLS for blood pressure per age.

5 Exercise #3.3

I preprocessed the data to calculate the ratio of damaged parts in each trial.

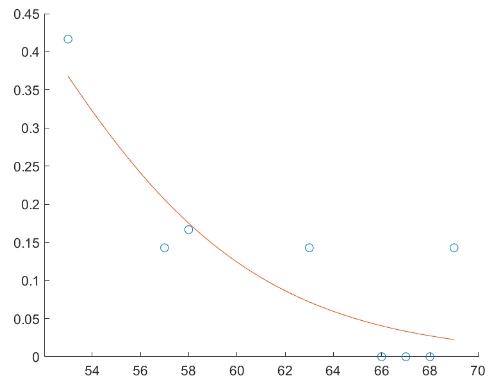

Figure 8: Logistic regression

```
%data
t = [53 57 58 63 66 67 67 67 68 69];
T = [ones(1, (size(t, 2))); t]
dam = [5 1 1 1 0 0 0 0 0 1];
udam = [7 6 5 6 8 8 7 6 5 6];
p_dam = (dam./(dam+udam))'
```

Let $K = 10$ and $b0, b1$ be the parameter to optimize so that

$$p(T) = \frac{1}{1 + e^{-(b0+b1x)}}$$

Let y be p_dam above indicating percent of damage for each trial. p be the vector with $i - th$ element valued $p(t_i)$, where t_i is the temperature at $i - th$ trial. p should be the function of $b0, b1$. Now, we want to minimize

$$\min_{b0,b1} \sum_{k=1}^{K} -y_k \ln p_k - (1 - y_k)\ln(1 - p_k)$$

which is the primal problem for logistic regression model.

(b)By using *glmfit* function in MATLAB, we obtain logistic regression model as following: $b0 = 10.1583, b1 = -0.2019$ so that

$$p(T) = \frac{1}{1 + e^{-(10.1583-0.2019T)}}$$

We have figure8 as follow of scattered data and logistic regression plot. x-axis is for the temperature and y-axis is for the probability of damage.

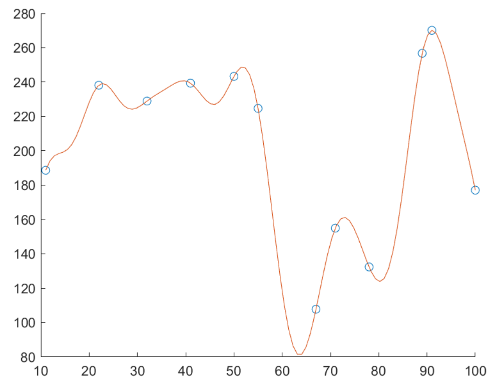

Figure 9: h = 5

6 Exercise #3.5

(a) By processing linear regression, the residual error is given by

$$r = 1.0e+03 * \begin{bmatrix} -0.0971 \\ 0.4565 \\ 0.1370 \\ 0.4373 \\ 0.2074 \\ -1.6348 \\ -0.6032 \\ 0.1605 \\ 0.5851 \\ -0.2483 \\ 0.0528 \\ 0.5468 \end{bmatrix}, \text{ with norm } ||r||_2 = 2.0588e+03$$

(c) Figure 9-11 are figures for the estimated function in feature space for given input x with $h = 5, 10, 15$ respectively. The dots represents the function values at sample data points. What we can observe is that the function gets more smooth(the Lipschitz coefficient for continuity decrease) as h increase.

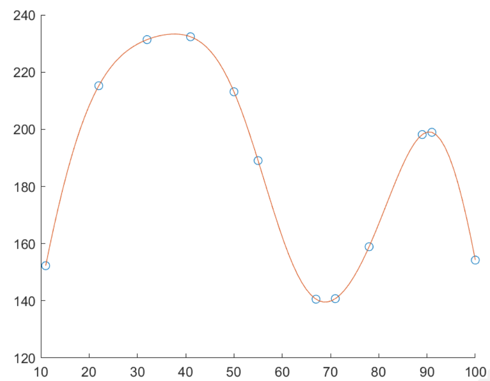

Figure 10: h = 10

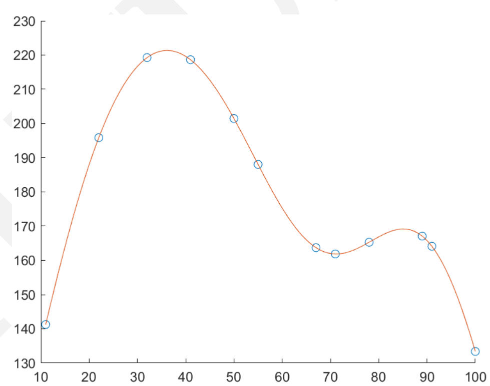

Figure 11: h = 15