Tarun Krishna
11593040
krishnatarun7@gmail.com

Homework Assignment 4
Machine Learning 1, 17/18

2017-10-11

1 First Exercise

1.1 First Subtask

$$
\begin{aligned}
f(x) &= 1 - x_1^2 - 2x_2^2 \\
g(x) &= x_1 + x_2 - 1 = 0
\end{aligned}
$$

$$L(x_1, x_2) = f(x) + \lambda(g(x))$$

$$
\begin{aligned}
\frac{\partial L}{\partial \lambda} &= x_1 + x_2 = 0 \\
\frac{\partial L}{\partial x_1} &= -2x_1 + \lambda = 0 \\
x_1 &= \frac{\lambda}{2} \\
\frac{\partial L}{\partial x_2} &= -4x_2 + \lambda = 0 \\
x_2 &= \frac{\lambda}{4}
\end{aligned}
$$

$$
\begin{aligned}
x_1 + x_2 &= 0 \\
\frac{\lambda}{2} + \frac{\lambda}{4} &= 1 \\
\lambda &= \frac{4}{3}
\end{aligned}
$$

$$
\begin{aligned}
x_1 &= \frac{2}{3} \\
x_2 &= \frac{1}{3}
\end{aligned}
$$

1.2 Second Subtask

$$
\begin{aligned}
f(x) &= 1 - x_1^2 - x_2^2 \\
g(x) &= x_1 + x_2 - 1 \geq 0
\end{aligned}
$$

$$g(x) \geq 0 \qquad \lambda \geq 0 \qquad \lambda g(x) = 0$$

$$
\begin{aligned}
L(x_1, x_2) &= f(x) + \lambda g(x) \\
&= 1 - x_1^2 - x_2^2 + \lambda(x_1 + x_2 - 1)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial L}{\partial \lambda} &= x_1 + x_2 - 1 = 0 \\
\frac{\partial L}{\partial x_1} &= -2x_1 + \lambda = 0 \\
x_1 &= \frac{\lambda}{2}
\end{aligned}
$$

$$\frac{\partial L}{\partial x_2} = -2x_2 + \lambda = 0$$
$$x_2 = \frac{\lambda}{2}$$

If $\lambda = 0$ then $g(x) > 0$ using complementary slack conditions

$$x_1 = 0 \qquad x_2 = 0$$
$$f(0,0) = 1$$

If $\lambda \neq 0$ then $g(x) = 0$ using complementary slack conditions

$$x_1 + x_2 = 1$$
$$\lambda = 1 \Rightarrow x_1 = \frac{1}{2}, x_2 = \frac{1}{2}$$
$$f(\frac{1}{2}, \frac{1}{2}) = \frac{1}{2}$$

Hence the maxima comes at $x_1 = 0$ and $x_2 = 0$

1.3 Third Subtask

$$f(x) = x_1 + 2x_2 - 2x_3$$
$$g(x) = x_1^2 + x_2^2 + x_3^2 - 1 = 0$$

$$L(x_1, x_2, x_3) = f(x) + \lambda g(x)$$
$$= x_1 + 2x_2 - 2x_3 + \lambda(x_1^2 + x_2^2 + x_3^2 - 1)$$

$$\frac{\partial L}{\partial \lambda} = x_1^2 + x_2^2 + x_3^2 - 1 = 0$$
$$\frac{\partial L}{\partial x_1} = 1 + 2\lambda x_1 = 0$$
$$x_1 = \frac{-1}{2\lambda}$$
$$\frac{\partial L}{\partial x_2} = 2 + 2\lambda x_2 = 0$$
$$x_2 = \frac{-1}{\lambda}$$
$$\frac{\partial L}{\partial x_3} = -2 + 2\lambda x_3 = 0$$
$$x_3 = \frac{1}{\lambda}$$

$$x_1^2 + x_2^2 + x_3^2 - 1 = 0$$
$$\frac{-1}{2\lambda}^2 + \frac{-1}{\lambda}^2 + \frac{1}{\lambda} = 0$$
$$\lambda = \pm\frac{3}{2}$$
$$\text{using } \lambda = \frac{3}{2}$$
$$x_1 = \frac{-1}{3} \qquad x_2 = \frac{-2}{3} \qquad x_3 = \frac{2}{3}$$
$$f(\frac{-1}{3}, \frac{-2}{3}, \frac{2}{3}) = -3$$
$$\text{using } \lambda = \frac{-3}{2}$$
$$x_1 = \frac{1}{3} \qquad x_2 = \frac{2}{3} \qquad x_3 = \frac{-2}{3}$$
$$f(\frac{1}{3}, \frac{2}{3}, \frac{-2}{3}) = 3 \leftarrow \text{max value}$$

1.4 Fourth Subtask

$$\begin{aligned}
f(x_1, x_2) &= 1 - x_1^2 - x_2^2 \\
g(x_1, x_2) &= -x_1 - x_2 + 1 \geq 0 \\
L(x_1, x_2) &= f(x) + \lambda g(x) \\
&= 1 - x_1^2 - x_2^2 + \lambda(x_1 - x_2 + 1)
\end{aligned}$$

$$g(x) \geq 0 \qquad \lambda \geq 0 \qquad \lambda g(x) = 0$$

$$\begin{aligned}
\frac{\partial L}{\partial \lambda} &= -x_1 - x_2 + 1 = 0 \\
\frac{\partial L}{\partial x_1} &= -2x_1 - \lambda = 0 \\
x_1 &= \frac{-\lambda}{2} \\
\frac{\partial L}{\partial x_2} &= -2x_2 - \lambda = 0 \\
x_2 &= \frac{-\lambda}{2}
\end{aligned}$$

Again using complimentary slack conditons $\lambda g(x) = 0$

if $\lambda = 0$

$$x_1 = 0, x_2 = 0 \Rightarrow f(0, 0) = 1$$

if $\lambda \neq 0$

$$-x_1 - x_2 + 1 + 1 = 0$$

$$\frac{\lambda}{2} + \frac{\lambda}{2} + 1 = 0 \leftarrow \lambda$$ negative conflict occurs

Hence maxima attained at $f(0, 0) = 1$

1.5 Fifth Subtask

Given, $Dose(D) = 6x^{\frac{2}{3}}y^{\frac{1}{2}}$ where cost of $x = 4$ euro per gram and $y = 3$ euro per gram. And Max budget is 7000 euro.

$$f(x,y) = 6x^{\frac{2}{3}}y^{\frac{1}{2}}$$
$$g(x,y) = -x - y + 7000 \geq 0$$

$$\begin{aligned} L(x,y) &= f(x,y) + \lambda g(x,y) \\ &= 6x^{\frac{2}{3}}y^{\frac{1}{2}} + \lambda(-x - y + 7000 \geq 0) \end{aligned}$$

$$\begin{aligned} \frac{\partial L}{\partial x} &= 4x^{\frac{-1}{3}}y^{\frac{1}{2}} - 4\lambda = 0 \\ \frac{y^{\frac{1}{2}}}{x^{\frac{1}{3}}} &= \lambda \\ \frac{\partial L}{\partial y} &= 3x^{\frac{2}{3}}y^{-1}2 - 3\lambda = 0 \\ \frac{x^{\frac{2}{3}}}{y^{\frac{1}{2}}} &= \lambda \end{aligned}$$

$$\begin{aligned} \frac{x^{\frac{2}{3}}}{y^{\frac{1}{2}}} &= \frac{y^{\frac{1}{2}}}{x^{\frac{1}{3}}} \\ x &= y \\ -4x - 3y + 7000 = 0 &\qquad -x - y + 7000 \geq 0 \\ -4x - 3x + 7000 &\geq 0 \\ x &= 1000 \\ y &= 1000 \\ Dose =& 6x^{\frac{2}{3}}y^{\frac{1}{2}} \\ &= 6000 \end{aligned}$$

2 Second Exercise

$$\min_{a,R,\xi} R^2 + C\sum_{i=1}^{N}\xi_i$$
$$s.t.\forall_i ||\mathbf{x}_i - \mathbf{a}||^2 \leq R^2 + \xi_i, \xi_i \geq 0$$

2.1 First Subtask

Primal Lagrangian is given by:

$$L(a,R,\xi,\alpha_i,\mu_i) = R^2 + C\sum_{i=1}^{N}\xi_i + \sum_{i=1}^{N=1}\alpha_i\big(||\mathbf{x}_i - \mathbf{a}||^2 - R^2 - \xi_i\big) - \sum_{i=1}^{N=1}\mu_i\xi_i$$

2.2 Second Subtask

Write down all KKT conditions:

$$L(a,R,\xi,\alpha_i,\mu_i) = R^2 + C\sum_{i=1}^{N}\xi_i + \sum_{i=1}^{N=1}\alpha_i\big(||\mathbf{x}_i - \mathbf{a}||^2 - R^2 - \xi_i\big) - \sum_{i=1}^{N=1}\mu_i\xi_i$$

(a)

$$\nabla_{R^2}L = 1 - \sum_{i=1}^{N}\alpha_i = 0 \Rightarrow \sum_{i=1}^{N}\alpha_i = 1$$

(b)

$$\begin{aligned}
\nabla_a L &= \nabla_a\Big(\sum_{i=1}^{N=1} \underbrace{\alpha_i\big(||\mathbf{x}_i - \mathbf{a}||^2\big)}_{\text{depends on }\mathbf{a}}\Big) = 0 \\
&= \nabla_a\Big(\sum_{i=1}^{N=1} \alpha_i(\mathbf{x}_i^T\mathbf{x}_i - 2\mathbf{x}_i^T a + a^T a)\Big) = 0 \\
\nabla_a L &= \sum_{i=1}^{N=1} \alpha_i(-2\mathbf{x}_i + 2\alpha_i) = 0 \\
\nabla_a L &= \sum_{i=1}^{N=1} \alpha_i a = \sum_{i=1}^{N=1} \alpha_i \mathbf{x}_i \Rightarrow a = \frac{\sum_{i=1}^{N=1} \alpha_i \mathbf{x}_i}{\sum_{i=1}^{N=1} \alpha_i} \Rightarrow a = \sum_{i=1}^{N=1} \alpha_i \mathbf{x}_i
\end{aligned}$$

(c)

$$\nabla_{\xi_i} L = C - \alpha_i - \mu_i = 0 \Rightarrow C = \alpha_i + \mu_i$$

(d)

$$\alpha_i\big(||\mathbf{x}_i - \mathbf{a}||^2 - R^2 - \xi_i\big) = 0$$

(e)

$$\mu_i \xi_i = 0$$

(f)

$$\mu_i \geq 0$$

(g)

$$\alpha i \geq 0$$

(h)

$$-\xi_i \leq 0$$

(i)

$$||\mathbf{x}_i - \mathbf{a}||^2 - R^2 - \xi_i \leq 0$$

2.3 Second Subtask

If x_i is inside the sphere $\Rightarrow ||\mathbf{x}_i - \mathbf{a}||^2 - R^2 - \xi_i \leq 0 \Rightarrow \alpha_i = 0 \Rightarrow \mu_i = C$(from (c))
If x_i is outside the sphere $\Rightarrow \xi_i > 0 \Rightarrow \mu_i = 0 \Rightarrow \alpha_i = C$
if x_i is outside or on the ball $\Rightarrow ||\mathbf{x}_i - \mathbf{a}||^2 - R^2 - \xi_i = 0 \Rightarrow \alpha_i \geq 0$
if x_i is inside or on the ball $\Rightarrow \xi_i = 0 \Rightarrow \mu_i \geq 0$

2.4 Second Subtask

$$L(a, R, \xi, \alpha_i, \mu_i) = R^2 + C\sum_{i=1}^{N} \xi_i + \sum_{i=1}^{N=1} \alpha_i\big(||\mathbf{x}_i - \mathbf{a}||^2 - R^2 - \xi_i\big) - \sum_{i=1}^{N=1} \mu_i \xi_i$$

Make it independent of a, R, ξ_i

$$\begin{aligned}
L(\alpha_i, \mu_i) &= R^2 + C\sum_{i=1}^{N} \xi_i + \sum_{i=1}^{N} \alpha_i(\mathbf{x}_i^T\mathbf{x}_i - 2\mathbf{x}_i^T a + a^T a - R^2 - \xi_i) - \sum_{i=1}^{N} \mu_i \xi_i \\
&= \big(R^2 - R^2\sum_{i=1}^{N} \alpha_i\big) + \sum_{i=1}^{N} \xi_i(C - \alpha_i - \mu_i) - 2\sum_{i=1}^{N}(\alpha_i \mathbf{x}_i^T)a + a^T a \sum_{i=1}^{N} \alpha_i + \sum_{i=1}^{N}
\end{aligned}$$

Intial two terms goes to zero using a, b, c

$$= -2a\sum_{i=1}^{N}(\alpha_i \mathbf{x}_i^T) + a^T a \sum_{i=1}^{N}\alpha_i + \sum_{i=1}^{N}\alpha_i \mathbf{x}_i^T \mathbf{x}_i$$

$\sum_{i=1}^{N}\alpha_i$ this term goes to 1 using (a), $\sum_{i=1}^{N}(\alpha_i \mathbf{x}_i^T)$ this can be written as a^T using (c)

$$\begin{aligned}
&= -2aa^T + a^T a + \sum_{i=1}^{N}\alpha_i \mathbf{x}_i^T \mathbf{x}_i \\
&= \sum_{i=1}^{N}\alpha_i \mathbf{x}_i^T \mathbf{x}_i - a^T a \\
&= \sum_{i=1}^{N}\alpha_i ||\mathbf{x}||^2 - \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j \mathbf{x}_i \mathbf{x}_j
\end{aligned}$$

so far we have gathered terms. Let's now kernalize it

$$\sum_{i=1}^{N}\alpha_i ||\mathbf{x}||^2 - \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j \mathbf{x}_i \mathbf{x}_j = \sum_{i=1}^{N}\alpha_i K(\mathbf{x}_i, \mathbf{x}_i) - \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i K(\mathbf{x}_i, \mathbf{x}_j)\alpha_j$$

the dual is:

$\arg\max_{\alpha} \sum_{i=1}^{N}\alpha_i K(\mathbf{x}_i, \mathbf{x}_i) - \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i K(\mathbf{x}_i, \mathbf{x}_j)\alpha_j$

$\alpha_i \in [0, C]\forall_i$

2.5 Fifth Subtask

So we have α^* as optimal value.

$$\begin{aligned}
\mu_i^* &= C - \alpha_i^* \\
a^* &= \sum_{i=1}^{N=1}\alpha_i^* \mathbf{x}_i \\
R^{2*} &= ||\mathbf{x}_i - \mathbf{a}^*||^2 \\
\xi_i &= \begin{cases} 0, & \alpha^* = 0 \\ ||\mathbf{x}_i - \mathbf{a}^*||^2 - R^{2*}, & \alpha_i^* = C \end{cases}
\end{aligned}$$

2.6 Sixth Subtask

A new test case is x_t is outlier when:

$$\begin{aligned}
||\mathbf{x}_t - \mathbf{a}^*||^2 &> R^{2*} \\
\mathbf{x}_t^T \mathbf{x}_t - 2\mathbf{x}_t^T a^* + a^{*T} a^* &> R^{2*} \\
K(x_t, x_t) - 2\sum_{i=1}^{N}\alpha_i^* \mathbf{x}_t^T \mathbf{x}_i + \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i^* K(\mathbf{x}_t, \mathbf{x}_i)\alpha_j^* &> R^{2*} \\
K(x_t, x_t) - 2\sum_{i=1}^{N}\alpha_i^* K(\mathbf{x}_t, \mathbf{x}_i) + \sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i^* K(\mathbf{x}_i, \mathbf{x}_j)\alpha_j^* &> R^{2*}
\end{aligned}$$

2.7 Seventh Subtask

When C goes to infinity all data cases will come inside the sphere. Also when C is zero, I guess R will be zero.

2.8 Eigth Subtask

With very small bandwith the solution will be overfitted.As nothing will be detected as outlier.
An RBF kernel will result in somekind of flexible boundaries. whereas linear kernel results in circular boundary.

2.9 Ninth Subtask

$$\min_{a,R,\xi} R^2 + C\sum_{i=1}^{N} \xi_i$$

In order to do that we replace it with soft margin. Which gives

$$y_i(||\mathbf{x}_i - \mathbf{a}||^2 - R^2) \geq 1 - \xi_i$$

Also $\xi_i \geq 0$

3 Third Exercise

3.1 First Subtask

$$x = \begin{bmatrix} 0.1 \\ 0.35 \end{bmatrix} \qquad W^{(1)} = \begin{bmatrix} w_1 & w_3 \\ w_2 & w_4 \end{bmatrix} = \begin{bmatrix} 0.4 & 0.65 \\ 0.2 & 0.01 \end{bmatrix} \qquad W^{(2)} = \begin{bmatrix} w_5 \\ w_6 \end{bmatrix} = \begin{bmatrix} 0.8 \\ 0.15 \end{bmatrix}$$

$$\begin{aligned}
Z^{(1)} &= (W^{(1)})^T X \\
\\
&= \begin{bmatrix} 0.4 & 0.65 \\ 0.2 & 0.01 \end{bmatrix} \begin{bmatrix} 0.1 \\ 0.35 \end{bmatrix} = \begin{bmatrix} 0.2675 \\ 0.0235 \end{bmatrix} \\
a^{(1)} &= \begin{bmatrix} 0.5664 \\ 0.5058 \end{bmatrix} \\
Z^{(2)} &= (W^{(2)})a^{(1)} \\
&= \begin{bmatrix} 0.8 & 0.15 \end{bmatrix} \begin{bmatrix} 0.5664 \\ 0.5058 \end{bmatrix} = \begin{bmatrix} 0.5289 \end{bmatrix} a^{(2)} = \begin{bmatrix} 0.6292 \end{bmatrix} \\
E &= -(tlny - (1-t)ln(1-y)) \qquad t = 1 \\
&= -ln(0.6292) \\
&= 0.4633
\end{aligned}$$

3.2 Second Subtask

$$\delta^{(2)} = -\frac{1}{a} \times \sigma^{'}(Z^{(2)})$$
$$= -(1-a) = -0.3708$$

$$\delta^{(1)} = W^{(2)}\delta^{2} \times \sigma^{'}(Z^{(1)})$$
$$= \begin{bmatrix} 0.8 \\ 0.15 \end{bmatrix} \begin{bmatrix} -0.3708 \end{bmatrix} \times \begin{bmatrix} a_1^{(1)}(1-a_1^{(1)}) \\ a_2^{(1)}(1-a_2^{(1)}) \end{bmatrix} \qquad \times \text{ is element wise multiplication}$$

$$= \begin{bmatrix} -0.29664 \\ -0.05562 \end{bmatrix} \times \begin{bmatrix} 0.2455 \\ 0.2499 \end{bmatrix}$$
$$= \begin{bmatrix} -0.0728 \\ -0.0138 \end{bmatrix}$$
$$\frac{\partial E}{\partial w^{(1)}} = a^{(1)}(\delta^{1})^{T}$$
$$= \begin{bmatrix} 0.1 \\ 0.35 \end{bmatrix} \begin{bmatrix} -0.0728 & -0.0138 \end{bmatrix} = \begin{bmatrix} -0.00728 & -0.00138 \\ -0.02548 & -0.00483 \end{bmatrix}$$
$$\frac{\partial E}{\partial w^{(2)}} = \begin{bmatrix} 0.5664 \\ 0.5056 \end{bmatrix} \begin{bmatrix} -0.3708 \end{bmatrix}$$
$$= \begin{bmatrix} -0.2097 \\ -0.1874 \end{bmatrix}$$
$$W^{(1)} = W^{(1)} - \eta(\frac{\partial E}{\partial W^{(1)}})$$
$$= \begin{bmatrix} 0.4 & 0.65 \\ 0.2 & 0.01 \end{bmatrix} - 0.05 \begin{bmatrix} -0.00728 & -0.00138 \\ -0.02548 & -0.00483 \end{bmatrix}$$
$$= \begin{bmatrix} 0.400036 & 0.650069 \\ 0.20127 & 0.1002415 \end{bmatrix}$$
$$W^{(2)} = W^{(2)} - \eta(\frac{\partial E}{\partial W^{(2)}})$$
$$= \begin{bmatrix} 0.8 \\ 0.15 \end{bmatrix} - 0.05 \begin{bmatrix} -0.2097 \\ -0.1874 \end{bmatrix}$$
$$= \begin{bmatrix} 0.81048 \\ 0.1593 \end{bmatrix}$$

W^1 is coressponding to first layer and W^2 next layer

3.3 Third subtask

$$Z^{(1)} = \begin{bmatrix} 0.4000364 & 0.650069 \\ 0.201274 & 0.1002418 \end{bmatrix} \begin{bmatrix} 0.1 \\ 0.35 \end{bmatrix} = \begin{bmatrix} 0.270279 \\ 0.05711925 \end{bmatrix}$$
$$a^{(1)} = \begin{bmatrix} 0.5671 \\ 0.5142 \end{bmatrix}$$
$$Z^{2} = \begin{bmatrix} 0.8, 0.15 \end{bmatrix} \begin{bmatrix} 0.5675 \\ 0.5142 \end{bmatrix} = \begin{bmatrix} 0.53722 \end{bmatrix}$$
$$a^{2} = \sigma(Z^{2}) = 0.6321$$
$$E = -ln(0.6321) \qquad \text{using cross entropy}$$
$$= 0.4587$$