Homework 3

Maxwell Aladago

March 28, 2018

Q2a

Given

$$
\begin{aligned}
&\arg\min_{w,b} \frac{1}{2}||w||^2 \\
s.t \qquad & y^{(i)}\left(w^T x^{(i)} + b\right) \geq 1, \forall i = 1, ..., m
\end{aligned} \tag{1}
$$

If

$$
\begin{aligned}
y &= [y^{(1)}, y^{(2)}, ..., y^{(m)}]^T \\
X &= [x^{(1)^T}, x^{(2)^T}, ..., x^{(m)^T}]
\end{aligned}
$$

Then Eqn 1 can be defined as

$$
\begin{aligned}
&\arg\min_{w,b} \frac{1}{2}||w||^2 + 0(b^2) \qquad s.t \qquad y\,(Xw + b) \geq 1 \\
&\equiv \arg\min_{w,b} \frac{1}{2} w^T w + 0(b^2) \qquad s.t \qquad y(Xw) + yb \geq 1 \\
&\equiv \arg\min_{w,b} \frac{1}{2} w^T \qquad s.t \qquad -[yX, \quad y][w, b]^T \leq -1 \\
&\equiv \arg\min_{z} \frac{1}{2} z^T z \qquad s.t \qquad -[yX, \quad y]z \leq -1
\end{aligned} \tag{2}
$$

Eqn 2 is now in Solver's format where

$x = z = [w, b]^T$
$P \in \mathcal{R}^{(n+1)\times(n+1)}$, n is the dimension of w
$\quad P[i, j] = 0, \quad \forall_{i \neq j}$
$\quad P[i, i] = 1$ for $i, j = 1, ..., n$
$\quad P[n+1, n+1] = 0$
$q = \{0\}^{n+1} \qquad s.t. q[i] = 0, \quad \forall_{i=1,...,n+1}$
$G \in \mathcal{R}^{m\times(n+1)} = -[y^T X, y]$
$h = \{-1\}^m, \quad h[i] = -1, \quad \forall_{i=1,...,m}$

The output of solver, solver's is x comprising the weights and bias. $x = x[0 : n] = w$ the weights and $x[n]$ = b, the bias.

Q2d

SVM primal with slack variable. Given

$$\begin{aligned} &\underset{\text{w, b, }\xi}{\arg\min}\, \frac{1}{2}||w||^2 + C\sum_{i=1}^{m}\xi_i \\ s.t \qquad & y^{(i)}\left(w^T x + b\right) \geq 1 - \xi_i, \quad \forall_i \\ & \xi_i \geq 0, \forall_i \end{aligned} \tag{3}$$

If

$$\begin{aligned} y &= [y^{(1)}, y^{(2)}, ..., y^{(m)}]^T \\ X &= [{x^{(1)}}^T, {x^{(2)}}^T, ..., {x^{(m)}}^T] \\ \xi &= [\xi_1, \xi_2, ..., \xi_m]^T \end{aligned}$$

Eqn 3 can be expressed as

$$\begin{aligned} &\underset{\text{w, b, }\xi}{\arg\min}\, \frac{1}{2}||w||^2 + C\xi \\ s.t \qquad & y(Xw) + yb + \xi \geq 1 \\ & \xi \geq 0 \\ &\equiv \underset{\text{w, b, }\xi}{\arg\min}\, \frac{1}{2}w^T w + 0(b^2 + \xi^T\xi) + 0(w + b) + C\xi \\ s.t \qquad & -\left(y(Xw) + yb + \xi\right) \leq -1 \\ & -\xi \leq 0 \end{aligned} \tag{4}$$

Eqn 4 can be expressed in solver's form

$$\begin{aligned} &\underset{x}{\arg\min}\, x^T P x + q^T x \\ s.t \qquad & Gx \leq h \end{aligned}$$

Where

$x = [w, b, \xi]^T$

$P \in \mathcal{R}^{(n+1+m)\times(n+1+m)}$, n = dimension of w, m = number of examples = dimension of ξ

$P[i,j] = 0, \quad \forall_{i\neq j} \quad \forall i > n \quad \forall j > n$

$P[i,i] = 1, \quad \forall_i \quad i = 1, ..., n$

$q \in \mathcal{R}^{n+1+m}, \quad q[i] = 0, q[j] = C \quad \forall i = 1, .., n+1, \quad \forall j = n+2, ..., n+m$

$let\ I =$ identity Matrix of dimension m

$z_w = \{0\}^{m\times n}, z_w[i,j] = 0,\ \forall i, j$

$z_b = \{0\}^m, zero_b[i] = 0,\ \forall i$

$$G \in \mathcal{R}^{2m\times(n+1+m)} = -\begin{bmatrix} y^T X & y & I \\ z_w & z_b & I \end{bmatrix}$$

$h \in \{-1, 0\}^{2m} \quad h[i] = -1, h[j] = 0 \quad \forall i = 1, ..., m, \quad \forall j = m+1, ..., 2m$

The output of solver, x has n + 1 + m values. $x[0:n] = w$, $x[n] = b$ and $x[n:n+m]$= the slack values

Q2f

The margins of $c = 100$ look exactly as before. This is because of the huge penalty for violating the margin constraint. There are violations of the margin constraint for $C = 0.1$. this is because the cost of violating the constraint is minimal.

Q2g

The dual formulation:

Given:

$$\begin{aligned} &\arg\max_{\alpha} \quad \sum_{i=1}^{m} \alpha_i - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y^{(i)} y^{(j)} k(x^{(i)}, x^{(j)}) \\ s.t \quad &\sum_{i=1}^{m} \alpha_i y^{(i)} = 0 \\ &0 \leq \alpha_i \leq C, \quad \forall_i \end{aligned} \tag{5}$$

The dual formulation in Eqn 5 is equivalent to

$$\begin{aligned} &\arg\min_{\alpha} \quad \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} \alpha_i \alpha_j y^{(i)} y^{(j)} k(x^{(i)}, x^{(j)}) - \sum_{i=1}^{m} \alpha_i \\ s.t \quad &\sum_{i=1}^{m} \alpha_i y^{(i)} = 0 \\ &0 \leq \alpha_i \leq C, \quad \forall_i \end{aligned} \tag{6}$$

If

$$\begin{aligned} \alpha &= [\alpha_1, \alpha_2, ..., \alpha_m]^T \\ y &= [y^{(1)}, y^{(m)}, ..., y^{(m)}]^T \\ K(x,x) &= \begin{bmatrix} k(x^{(1)}, x^{(1)} & k(x^{(1)}, x^{(2)} & \cdots & k(x^{(1)}, x^{(m)} \\ \vdots & & & \vdots \\ k(x^{(m)}, x^{(1)} & k(x^{(m)}, x^{(2)} & \cdots & k(x^{(m)}, x^{(m)} \end{bmatrix} \end{aligned}$$

Eqn 6 can be expressed as

$$\begin{aligned} &\arg\min_{\alpha} \quad \frac{1}{2}\alpha^T y y^T K(x,x)\alpha - \alpha \\ s.t \quad &y^T\alpha = 0 \\ &-\alpha \leq 0 \\ &\alpha \leq C \end{aligned} \tag{7}$$

Eqn 7 can be expressed in solvers form

$$
\begin{aligned}
&\arg\min_x x^T P x + q^T x \\
s.t \qquad & Gx \leq h \\
& Ax = b
\end{aligned}
$$

Where

$$
\begin{aligned}
x &= \alpha \\
P &= yy^T K(x, x) \\
q &= \{-1\}^m, \quad q[i] = -1 \quad \forall i \\
let\, g_0 &= \{-1, 0\}^{m \times m}, \quad g_0[i, j] = 0 \quad \forall_{i \neq j}, \quad g_0[i, i] = -1 \quad \forall i \\
g_c &= 0, C^{m \times m}, \quad g_c[i, j] = 0 \quad \forall_{i \neq j}, \quad g_c[i, i] = C \quad \forall i \\
G &= \begin{bmatrix} g_0 \\ g_c \end{bmatrix} \\
h \ &= \{0, C\}^{2m}, \quad h[i] = 0, \quad h[j] = C \quad \forall i = 1, ..., m, \quad \forall j = m + 1, ..., 2m \\
A &\in \mathcal{R}^{1 \times m} = y^T. \ A \text{ has only one row but m columns} \\
b &= 0
\end{aligned}
$$

Q2h

The linear kernel produced 3 support vectors. The polynomial case obtained 3 support vectors too. Yes, the decision boundary is bent. It looked superficially straight because the features are linearly separable.