ML Assignment 1

Ankit Kumar 170121
Chaitanya Pathare 19111061
Hemant Kumar 170297
Rahul Ninaji Pakhare 19111068
Raunak Kumar 19111069

1.

Given equation

$$argmin_{w\epsilon R^d}\frac{1}{2}||\mathbf{w}||_2^2 + C\Sigma_{i=1}^{n}([1 - y^i < w, x^i >]_+)^2 \quad (P1)$$

rewritten as new optimization problem

$$argmin_{w\epsilon R^d}\frac{1}{2}||\mathbf{w}||_2^2 + C\Sigma_{i=1}^{n}\xi_i^2$$

$$s.t. y^i < w, x^i > \geqslant 1 - \xi_i, for \quad all \quad i\epsilon[n] \quad (P2)$$

Introducing dual variable α_i for each of the constraints in (P2)

$$\therefore L(w, \xi, \alpha) = \frac{1}{2}||\mathbf{w}||_2^2 + C\Sigma_{i=1}^{n}\xi_i^2 + \Sigma_{i=1}^{n}\alpha_i(1 - \xi_i - y^i < w, x^i >) \qquad (P3)$$

2. To get the dual problem from the above equation (P3) put $\frac{\eth L}{\eth w} = 0 and \frac{\eth L}{\eth \xi} = 0$

$$\frac{\eth L}{\eth w} = ||\mathbf{w}|| - \Sigma_{i=1}^{n} \alpha_i x_i = 0$$

$$or \quad ||\mathbf{w}|| = \Sigma_{i=1}^{n} \alpha_i x^i$$

and

$$\frac{\eth L}{\eth \xi} = 2C\xi_i - \Sigma_{i=1}^{n} \alpha_i = 0$$

$$or \quad \xi_i = \frac{1}{2C} \Sigma_{i=1}^{n} \alpha_i$$

Optimization problem comes down to

$$= \frac{1}{2} \Sigma_{i=1}^{n} \Sigma_{j=1}^{n} (\alpha_i \alpha_j y^i y^j < x^i x^j >) + \frac{1}{4C} \Sigma_{i=1}^{n} \alpha_i^2 + \Sigma_{i=1}^{n} \alpha_i - \frac{1}{2C} \Sigma_{i=1}^{n} \alpha_i^2$$

$$- \Sigma_{i=1}^{n} \Sigma_{j=1}^{n} (\alpha_i x^i y^i)(\alpha_j x^j y^j))$$

$$= argmin(\Sigma_{i=1}^{n} \alpha_i - \frac{1}{4C} (\Sigma_{i=1}^{n} \alpha_i^2) - \frac{1}{2} \Sigma_{i=1}^{n} \Sigma_{j=1}^{n} (\alpha_i \alpha_j y^i y^j < x^i x^j >)) \quad (D2)$$

$$All \quad above \quad \mathbf{w} \epsilon R^{d+1}, \quad \mathbf{x} \epsilon R^{d+1}$$

3.
a. Stochastic coordinate descent
Given,

$$P1 \quad f = \quad argmin_{w \epsilon R^d} \frac{1}{2}||\mathbf{w}||_2^2 + C\Sigma_{i=1}^n([1 - y^i < w, x^i >]_+)^2$$

now get gradient for P1

$$\nabla f = ||\mathbf{w}|| + 2C\Sigma_{i=1}^n([1 - y^i < w, x^i >]_+)(-y^i x^i)$$

Note: $[1 - y^i < w, x^i >]_+ = \{(1 - y^i < w, x^i >) \quad if \quad y^i < w, x^i >< 1$

$$0 \quad if \quad y^i < w, x^i >\geqslant 1\}$$

iterate till the time end for the X

do 1.find gradient for each data point in the given Batch set B
2.update w for these $w = w - \eta 2C\Sigma_{i=x}^{x+B}([1 - y^i < w, x^i >]_+)(-y^i x^i)$

c. Method used for D2 minimization

$$D2 \qquad => argmin(\Sigma_{i=1}^n \alpha_i - \frac{1}{4C}(\Sigma_{i=1}^n \alpha_i^2) - \frac{1}{2}\Sigma_{i=1}^n \Sigma_{j=1}^n(\alpha_i \alpha_j y^i y^j < x^i x^j >))$$

concentrate on α_i

$$argmin(\alpha_i - \frac{1}{4C}(\alpha_i^2) - (\Sigma_{i \neq j}\alpha_j^2) - \frac{1}{2}(\alpha_i^2||x^i||^2) - \alpha_i y^i \Sigma_{i \neq j}(\alpha_j y^j < x^i x^j >))$$

$$Let \quad x = \alpha_i, \quad q = ||x^i||^2, \quad p = y^i \Sigma_{i \neq j}(\alpha_j y^j < x^i x^j >), \quad r = (\Sigma_{i \neq j}\alpha_j^2)$$

$$\therefore argmin(x^2(\frac{1}{2}q + \frac{1}{4C}) - x(1 - p) + r)$$

$$Let \quad \frac{q^0}{2} = (\frac{q}{2} + \frac{1}{4C}) \qquad \therefore argmin(\frac{q^0}{2}x^2 - x(1 - p) + r)$$

minimum at $x^m = \frac{(1-p)}{q^0}$
If $x^m \epsilon[0, \propto]$ then x^m is solution
elseif $x^m < 0$ solution is 0
else solution is $\propto$ or Not defined.
Note:

$$p = y^i \Sigma_{i \neq j}(\alpha_j y^j < x^i x^j >)$$

$$= w^T \mathbf{x}^i - \alpha_i y^i q^0$$

Algorithm minimize:

$$intialize \quad \alpha^T = \{0, 0, ...0\}_{1Xn} \quad w = \{0, 0, ...0\}_{1Xn}$$

$$w = \Sigma_{i=1}^{n} \alpha_i y_i x^i$$

Note: We have appended one extra dimension to our data i.e if X = $\begin{bmatrix} x_{11} & x_{12} & x_{13} & . & x_{1n} \\ x_{21} & . & . & . & x_{2n} \\ . & . & . & . & . \\ . & . & . & . & . \\ x_{n1} & . & . & . & x_{nn} \end{bmatrix}$

and then transformed X $\rightarrow X^0 = \begin{bmatrix} x_{11} & x_{12} & x_{13} & . & x_{1n} & 1 \\ x_{21} & . & . & . & x_{2n} & 1 \\ . & . & . & . & . & 1 \\ . & . & . & . & . & 1 \\ . & . & . & . & . & 1 \\ x_{n1} & . & . & . & x_{nn} & 1 \end{bmatrix}$

so that w[d-1] =b , d is dimension of each data point in X^0
iterate till time does not end, for each data point in X^0

Do - 1. Calculate α_i if $\alpha_i \geqslant 0$ then $\alpha[i] = \alpha_i$
else $\alpha[i] = 0$

2. Update w $w = \Sigma_{i=1}^{n} \alpha_i y_i x^i$

w will converge after sufficient number of iteration.