Assignment of chapter 5

Chengzhang Yang 16337272 Intelligent Science and Technology

4

(a)

Denotes r as the distance between x_a and the hyperplane. Decompose x_a into two vectors x and $\frac{w}{||w||}r$ where x locates on the hyperplane and r is parallel to w.

Then

$$\begin{aligned}|g(x_a)| &= |w^t x_a + w_0| \\ &= |w^t x + r||w|| + w_0| \\ &= r||w||\end{aligned}$$

Thus

$$r = \frac{|g(x_a)|}{||w||}$$

Suppose there is another point x' on the hyperplane.

Since $x - x'$ is an nonempty vector on hyperplane, we have

$$w^t(x - x') = w^t x + w_0 - (w^t x' + w_0) = 0$$

In other words, $x - x'$ is orthorgonal to w

Therefore

$$\begin{aligned}||x_a - x'||^2 = ||x - x'||^2 + r^2 \\ ||x_a - x'||^2 \geq ||x - x'||^2\end{aligned}$$

(b)

$$\begin{aligned}x_p = x &= \begin{cases} x_a - \frac{|g(x_a)|}{||w||}w, g(x_a) > 0 \\ x_a + \frac{|g(x_a)|}{||w||}w, g(x_a) < 0 \end{cases} \\ &= x_a - \frac{g(x_a)}{||w||^2}w\end{aligned}$$

14

(a)

$$\frac{\partial J_s(a)}{\partial a_j} = 2\sum_{i=1}^{n}(a^t y_i - b_i)y_{ij}$$
$$\frac{\partial J_s^2(a)}{\partial a_i a_j} = 2\sum_{i=1}^{n} y_{ij}^2$$
$$H = \begin{bmatrix} 12 & -2 & 12 \\ -2 & 70 & 72 \\ 12 & 72 & 288 \end{bmatrix}$$

(b)

Recall

$$a_{k+1} \approx a_k - \eta(k)\nabla J(a_k)$$

Consider second order Taylor expansion

$$J(a) \approx J(a_k) + \nabla J^t(a - a_k) + \frac{1}{2}(a - a_k)^t H(a - a_k)$$
$$J(a_{k+1}) \approx J(a_k) - \eta(k)||\nabla J||^2 + \frac{1}{2}\eta^2(k)\nabla J^t H \nabla J$$

Compute the derivative of the equation above, we know that $J(a_{k+1})$ achieve its minimum when

$$\eta(k) = \frac{||\nabla J||^2}{\nabla J^t H \nabla J}$$

Since α is unknown, we could not explicitly give the value of η. Nevertheless, we can estimate the range of $\eta(k)$. With the help of eigenvalue, we know that

$$\frac{1}{\lambda_1} = \frac{v_1^2 + v_2^2 + \ldots + v_n^2}{\lambda_1 v_1^2 + \lambda_1 v_2^2 + \ldots + \lambda_1 v_n^2} \le \frac{v_1^2 + v_2^2 + \ldots + v_n^2}{\lambda_1 v_1^2 + \lambda_2 v_2^2 + \ldots + \lambda_n v_n^2} \le \frac{v_1^2 + v_2^2 + \ldots + v_n^2}{\lambda_n v_1^2 + \lambda_n v_2^2 + \ldots + \lambda_n v_n^2} = \frac{1}{\lambda_n}$$

where λ_1 and λ_n refer to the largest and smallest eigenvalue of H. Therefore, the range of η is [0.006452, 0.1846].

21

Separate Y into four blocks.

$$Y = \begin{bmatrix} \mathbf{1}_1 & X_1 \\ -\mathbf{1}_2 & X_2 \end{bmatrix}$$

$$\because T^T Y a = Y^T b$$

$$\therefore \begin{bmatrix} \mathbf{1}_1^T & -\mathbf{1}_2^T \\ -X_1 & X_2^T \end{bmatrix} \begin{bmatrix} \mathbf{1}_1 & X_1 \\ -\mathbf{1}_2 & X_2 \end{bmatrix} a = \begin{bmatrix} \mathbf{1}_1 & X_1 \\ -\mathbf{1}_2 & X_2 \end{bmatrix} \begin{bmatrix} \frac{n}{n_1}\mathbf{1}_1 \\ \frac{n}{n_2}\mathbf{1}_2 \end{bmatrix}$$
$$\begin{bmatrix} n & (n_1 m_1 + n_2 m_2)^T \\ n_1 m_1 + n_2 m_2 & X_1^T X_1 + X_2^T X_2 \end{bmatrix} a = \begin{bmatrix} 0 \\ n(m_1 - m_2) \end{bmatrix} \tag{1}$$

Define

$$
\begin{aligned}
m_i &= \frac{1}{n_i}\sum_{x\in D_i} x \\
S_W &= \sum_{i=1}^{2}\sum_{x\in D_i}(x-m_i)(x-m_i)^T \\
&= \sum_{i=1}^{2}\sum_{x\in D_i}[(x-m_i)x^T-(x-m_i)m_i^T] \\
&= \sum_{i=1}^{2}\sum_{x\in D_i}[xx^T-2\frac{1}{n_i}\sum y\in D_i yx^T+m_im_i^T] \\
&= \sum_{i=1}^{2}[\sum_{j=1}^{n_i}x_jx_j^T-2\frac{1}{n_i}\sum_{k=1}^{n_i}x_k\sum_{l=1}^{n_i}x_k^T+n_im_im_i^T] \\
&= \sum_{i=1}^{2}[\sum_{j=1}^{2}x_jx_j^T-n_im_im_i^T] \\
&= X_1^TX_1+X_2^TX_2-n_1m_1m_1^T-n_2m_2m_2^T
\end{aligned}
$$

Equation(1) can be written as

$$
\begin{bmatrix} n & (n_1m_1+n_2m_2)^T \\ n_1m_1+n_2m_2 & S_W+n_1m_1m_1^T+n_2m_2m_2^T \end{bmatrix} a = \begin{bmatrix} 0 \\ n(m_1-m_2) \end{bmatrix}
$$

Solving equation set above, we have

$$
\begin{aligned}
(\frac{n_1n_2(m_1m_1^T+m_2m_2^T-2m_1m_2^T)}{n^2}+\frac{1}{n}S_W)w = m_1-m_2 \\
(\frac{n_1n_2(m_1-m_2)(m_1-m_2)^T}{n^2}+\frac{1}{n}S_W)w = m_1-m_2
\end{aligned} \tag{2}
$$

Since $(m_1-m_2)(m_1-m_2)^Tw$ is just linear combination of m_1-m_2 with different scalar, we can denote

$$
(1-\alpha)(m_1-m_2) = \frac{n_1n_2(m_1-m_2)(m_1-m_2)^T}{n^2}w
$$

Thus

$$
w = \alpha nS_W^{-1} \tag{3}
$$

Substitute w in equation(3) for equation(2) we have

$$
\alpha = (1+\frac{n_1n_2}{n}(m_1-m_2)^TS_W^{-1}(m_1-m_2))^{-1}
$$

28

(1) When samples are linear separable

$$
\exists a_0 \quad a_o^Ty_k = b_k, 1\le k\le n
$$

$$
0\le \min\{\tau:\tau\ge, a_i^Ty_i+\tau\ge b_i\} = \min\{\tau:\tau\ge, a_k^Ty_k+\tau\ge b_k\} = 0
$$

$$
\begin{aligned}
&\because \quad J_\tau(a)\ge 0, J_\tau(a_0)=0 \\
&\therefore \arg\max_{\alpha} J_\tau(a) = J_\tau(a_0)
\end{aligned}
$$

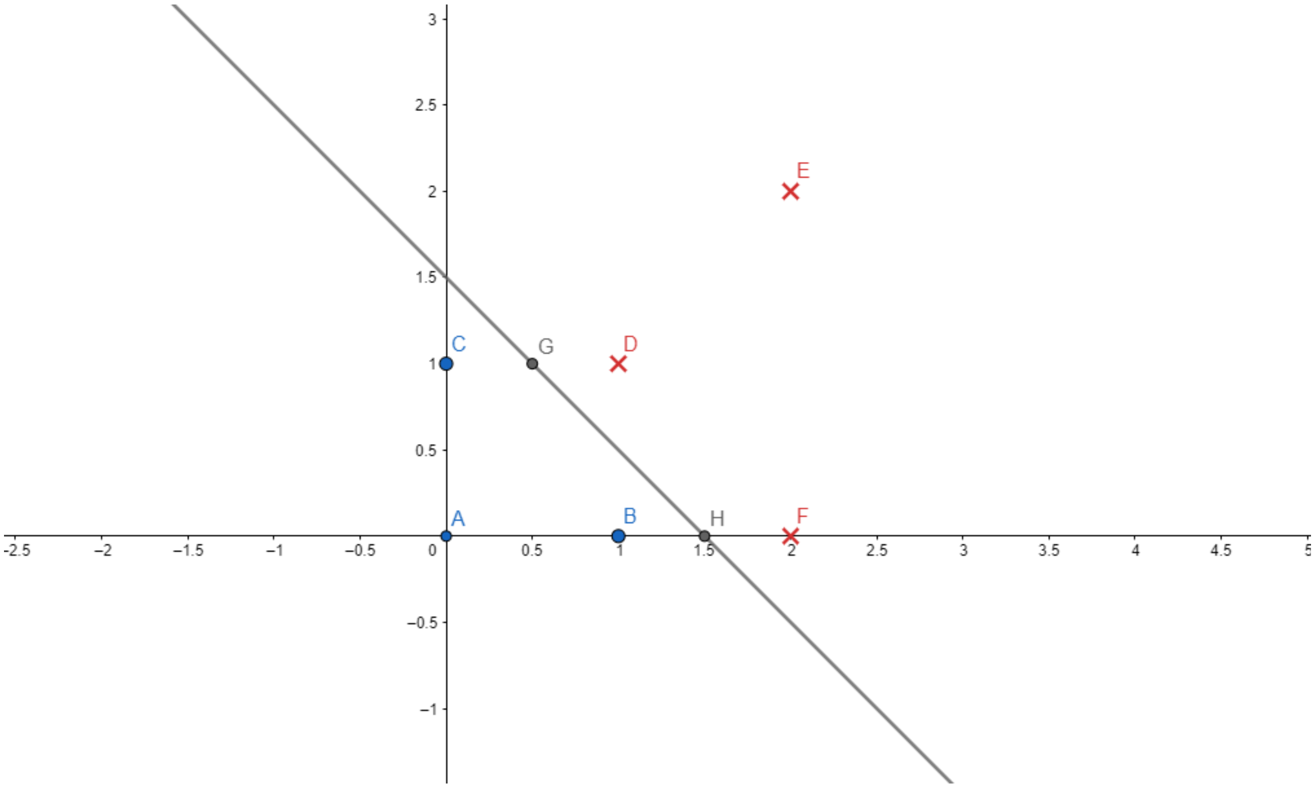
(2) When samples are not linear separable

$$\forall a, \quad a_0^T y_i < b_i, 1 \le i \le n$$

$$\begin{aligned}
\min_{\tau,a}\{\tau : \tau \ge 0, a^T y_i + \tau \ge b_i\} &= \min_{\tau,a}\{\tau : \tau \ge 0, \tau > b_i - a^T y_i\} \\
&= \min_{\tau,a}\{\tau : \tau \ge 0, \tau > \arg\max_a\{b_i - a^T y_i\}\} \\
&= \min_a\{\max_{a^T y_i \le b_i}\{b_i - a^T y_i\}\} \\
&= \min_a J_\tau(a)
\end{aligned}$$

32

(a)



$w = [3, -2, -2]^T$

(b)

point B, C, D, F

(c)

The dual problem is

$$maximize \;\; L_D(w,b,a) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n} \alpha_i\alpha_j y_i y_j x_i^T x_j$$

$$s.t. \;\; \alpha_i \ge 0 \;\; and \;\; \sum_{i=1}^{n} \alpha_i y_i = 0$$

Then

$$L_D = 2(a_1 + a_2 + a_3) - \frac{1}{2}\sum_{i=1}^{5}\sum_{j=1}^{5}\alpha_i\alpha_j y_i y_j x_i^T x_j$$

$$+\sum_{i=1}^{5}\alpha_k(a_1 + a_2 + a_3 - a_4 - a_5 - a_6)y_k x_6^T x_k$$

$$-\frac{1}{2}(a_1 + a_2 + a_3 - a_4 - a_5)^2 x_6^T x_6$$

Compute partial derivative of $a_1, a_2, a_3, \ldots, a_5$, we get

$$\begin{bmatrix} 1 & 2 & 2 & 0 & -1 \\ 2 & 5 & 3 & 1 & -1 \\ 2 & 3 & 5 & -1 & -3 \\ 0 & -1 & 1 & -1 & -1 \\ 1 & 1 & 3 & -1 & -2 \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 2 \\ 0 \\ 0 \end{bmatrix}$$

Simplify matrix of the right side

$$\begin{bmatrix} 1 & 2 & 2 & 0 & -1 \\ 0 & 1 & -1 & 1 & 1 \\ 0 & 1 & -1 & 1 & 1 \\ 0 & 1 & -1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} a' = \begin{bmatrix} 2 \\ -2 \\ 2 \\ 0 \\ 0 \end{bmatrix}$$

Obviously, this matrix is unsolvable. The optimal solution is on the boundary of the domain.

We can merely satisfy only one condition among equation (2)~(4).

(i) Assume that equation (2) holds

We can conclude that $a_3' > 0$ if all element of a' is greater or equal to zero.

However, in the condition above, we could not find a solution.

(ii) Assume that equation (3) holds

We can get

$$\begin{cases} a = [4 \quad 0 \quad 0 \quad 0 \quad 2 \quad 2]^T, \ L(a) = 4 \\ a = [2 \quad 0 \quad 0 \quad 2 \quad 0 \quad 0]^T, \ L(\alpha) = 2 \end{cases}$$

(iii) Assume that equation (4) holds

We can get

$$a = [2 \quad 0 \quad 0 \quad 0 \quad 0 \quad 2]^T, \ L(a) = 2$$

In summary, $a = [4 \quad 0 \quad 0 \quad 0 \quad 2 \quad 2]^T$

$$w = \sum_{i=1}^{6}\alpha_i y_i x_i = [2 \ \ 2]^T$$

Using point A, we have

$$b = \frac{1}{y_1} - w^T x_i = -3$$

Therefore, $w = [3, -2, -2]^T$, which is the identical solution of **(a)**