CSCI567TA3

Youzhi Qu USCID: 5045807938

November 2019

1 Principle Component Analysis

$L = \boldsymbol{v}_{M+1}^T \boldsymbol{X}^T \boldsymbol{X} \boldsymbol{v}_{M+1} - \alpha(\boldsymbol{v}_{M+1}^T \boldsymbol{v}_{M+1} - 1) - \sum_{i=1}^{M} \gamma_i \boldsymbol{v}_{M+1}^T \boldsymbol{v}_i$

subject to $\alpha \geqslant 0$, $\gamma_i \geqslant 0$, $\forall i$

$\frac{\partial L}{\partial \boldsymbol{v}} = 2\boldsymbol{X}^T \boldsymbol{X} \boldsymbol{v}_{M+1} - 2\alpha \boldsymbol{v}_{M+1} - \sum_{i=1}^{M} \gamma_i \boldsymbol{v}_i = 0(1)$

left multiply $\boldsymbol{v}_j^T, j \in [1, \cdots, M]$ to the equation can obtain:

$2\boldsymbol{v}_j^T \boldsymbol{X}^T \boldsymbol{X} \boldsymbol{v}_{M+1} - 2\alpha \boldsymbol{v}_j^T \boldsymbol{v}_{M+1} - \boldsymbol{v}_j^T \sum_{i=1}^{M} \gamma_i \boldsymbol{v}_i = 0$

since $\boldsymbol{v}_j^T \sum_{i=1}^{M} \gamma_i \boldsymbol{v}_i = \gamma_j \boldsymbol{v}_j^T \boldsymbol{v}_j = \gamma_j$ and $\boldsymbol{v}_j^T \boldsymbol{v}_{M+1} = 0$

$2(\boldsymbol{X}^T \boldsymbol{X} \boldsymbol{v}_j)^T \boldsymbol{v}_{M+1} - 0 - \gamma_j = 0$

$2\lambda_j \boldsymbol{v}_j^T \boldsymbol{v}_{M+1} - \gamma_j = 0$

we can get $\gamma_j = 0, j \in [1, \cdots, M]$

$\sum_{i=1}^{M} \gamma_i \boldsymbol{v}_i = 0(2)$

according to (1)(2),we can get $\boldsymbol{v}_{M+1}$ is an eigenvector of $\boldsymbol{X}^T \boldsymbol{X}$

$(\boldsymbol{X}^T \boldsymbol{X}) \boldsymbol{v}_{M+1} = \alpha \boldsymbol{v}_{M+1}(3)$

then:

$\boldsymbol{v}_{M+1}^T (\boldsymbol{X}^T \boldsymbol{X}) \boldsymbol{v}_{M+1} = \boldsymbol{v}_{M+1}^T \alpha \boldsymbol{v}_{M+1}$

$\boldsymbol{v}_{M+1}^T (\boldsymbol{X}^T \boldsymbol{X}) \boldsymbol{v}_{M+1} = \alpha$

when the $\boldsymbol{v}_{M+1}$ is the eigenvector with λ_{M+1}, quantity in $\boldsymbol{v}_{M+1}^T \boldsymbol{X}^T \boldsymbol{X} \boldsymbol{v}_{M+1}$ is maximized

2 Support Vector Regression

2.1

$\min_{\boldsymbol{w}, b} \frac{1}{2} \|\boldsymbol{w}\|_2^2 + C \sum_{n=1}^{N} (\xi_n + \xi_n')$

subject to

$$\begin{aligned}
&y_n - f(x_n) - \epsilon - \xi_n \le 0 \quad &, \quad \forall n \\
&f(x_n) - y_n - \epsilon - \xi_n' \le 0 \quad &, \quad \forall n \\
&-\xi_n \le 0 \quad &, \quad \forall n \\
&-\xi_n' \le 0 \quad &, \quad \forall n \\
&f(x_n) = \boldsymbol{w}^T \phi(\boldsymbol{x}_n) + b &
\end{aligned}$$

2.2

Lagrangian

$L = \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C\sum_{n=1}^N(\xi_n+\xi'_n) + \sum_{n=1}^N \alpha_n(y_n - f(x_n) - \epsilon - \xi_n) + \sum_{n=1}^N \alpha'_n(f(x_n) - y_n - \epsilon - \xi'_n) - \sum_{n=1}^N \beta_n\xi_n - \sum_{n=1}^N \beta'_n\xi'_n$

subject to: $\alpha_n \geq 0, \alpha'_n \geq 0, \beta_n \geq 0, \beta'_n \geq 0\ \forall n$

then:

$$\frac{\partial L}{\partial \boldsymbol{w}} = \boldsymbol{w} - \sum_{n=1}^N \alpha_n\Phi(x_n) + \sum_{n=1}^N \alpha'_n\Phi(x_n) = 0 \quad , \quad \boldsymbol{w} = \sum_{n=1}^N(\alpha_n - \alpha'_n)\Phi(x_n)$$

$$\frac{\partial L}{\partial b} = \sum_{n=1}^N(\alpha'_n - \alpha_n) = 0$$

$$\frac{\partial L}{\partial \xi_n} = C - \alpha_n - \beta_n = 0 \quad , \quad \forall n$$

$$\frac{\partial L}{\partial \xi'_n} = C - \alpha'_n - \beta'_n = 0 \quad , \quad \forall n$$

simplify expression:

$$\max_{\{\alpha_n\},\{\alpha'_n\}} -\frac{1}{2}\sum_{n,m=1}^N(\alpha_n - \alpha'_n)(\alpha_m - \alpha'_m)\Phi(x_n)^T\Phi(x_m) - \epsilon\sum_{n=1}^N(\alpha_n + \alpha'_n)$$

subject to $\sum_{n=1}^N(\alpha_n - \alpha'_n) = 0$ and $\alpha_n \geq 0, \alpha'_n \geq 0, \forall n$

3 Support Vector Machine

3.1

primal formulation:

$$\min_{\boldsymbol{w},b} \frac{1}{2}\|\boldsymbol{w}\|_2^2$$

subject to $y_n[\boldsymbol{w}^T\Phi(\boldsymbol{x}_n) + b] \geq 1, \forall n$

since:

$$\Phi(\boldsymbol{x}_1) = [1,0]^T \qquad y_1 = 1$$

$$\Phi(\boldsymbol{x}_2) = [0,1]^T \qquad y_2 = -1$$

$$\Phi(\boldsymbol{x}_3) = [-1,0]^T \qquad y_3 = 1$$

we can get:

$$\min_{\boldsymbol{w},b} \frac{1}{2}(w_1^2 + w_2^2)$$

subject to

$$w_1 + b \geq 1$$

$$-(w_2 + b) \geq 1$$

$$-w_1 + b \geq 1$$

dual formulation:

$$\max_{\{\alpha_n\}} \sum_n \alpha_n - \frac{1}{2}\sum_{m,n} y_m y_n \alpha_m \alpha_n k(\boldsymbol{x}_m, \boldsymbol{x}_n)$$

subject to:

$\sum_n \alpha_n y_n = 0$

$\alpha_n \geq 0$, $\forall n$

since:

$k(x_1, x_1) \quad \Phi(\boldsymbol{x}_1)^T \Phi(\boldsymbol{x}_1) = 1$

$k(x_1, x_2) \quad \Phi(\boldsymbol{x}_1)^T \Phi(\boldsymbol{x}_2) = 0$

$k(x_1, x_3) \quad \Phi(\boldsymbol{x}_1)^T \Phi(\boldsymbol{x}_3) = -1$

$k(x_2, x_2) \quad \Phi(\boldsymbol{x}_2)^T \Phi(\boldsymbol{x}_2) = 1$

$k(x_2, x_3) \quad \Phi(\boldsymbol{x}_2)^T \Phi(\boldsymbol{x}_3) = 0$

$k(x_3, x_3) \quad \Phi(\boldsymbol{x}_3)^T \Phi(\boldsymbol{x}_3) = 1$

then:

$y_1 y_1 \alpha_1 \alpha_1 k(x_1, x_1) = \alpha_1^2$

$y_1 y_2 \alpha_1 \alpha_2 k(x_1, x_2) = 0$

$y_1 y_3 \alpha_1 \alpha_3 k(x_1, x_3) = -\alpha_1 \alpha_3$

$y_2 y_1 \alpha_2 \alpha_1 k(x_2, x_1) = 0$

$y_2 y_2 \alpha_2 \alpha_2 k(x_2, x_2) = \alpha_2^2$

$y_2 y_3 \alpha_2 \alpha_3 k(x_2, x_3) = 0$

$y_3 y_1 \alpha_3 \alpha_1 k(x_3, x_1) = -\alpha_3 \alpha_1$

$y_3 y_2 \alpha_3 \alpha_2 k(x_3, x_2) = 0$

$y_3 y_3 \alpha_3 \alpha_3 k(x_3, x_3) = \alpha_3^2$

we can get:

$\max\limits_{\{\alpha_n\}} (\alpha_1 + \alpha_2 + \alpha_3) - \frac{1}{2}(\alpha_1^2 - 2\alpha_1\alpha_3 + \alpha_2^2 + \alpha_3^3)$

subject to:

$\alpha_1 - \alpha_2 + \alpha_3 = 0$

$\alpha_n \geq 0$ for $n = 1, \cdots, 3$

3.2

since $\alpha_2 = \alpha 1 + \alpha 3$ and dual formulation, we can get:

$\max\limits_{\{\alpha_n\}} 2\alpha_1 + 2\alpha_3 - \alpha_1^2 - \alpha_3^2$

subject to

$\alpha_1 \geq 0$ and $\alpha_3 \geq 0$

take derivative with respect to α_1, α_3 and set them to 0.

we can get:

$2 - 2\alpha_1 = 0$

$2 - 2\alpha_3 = 0$

then $\alpha_1 = 1, \alpha_2 = 2, \alpha_3 = 1$

$\boldsymbol{w}^* = \sum_n \alpha_n^* y_n \Phi(\boldsymbol{x}_n) = [0, -2]^T$

$b^* = y_n - \boldsymbol{w}^* \Phi(\boldsymbol{x}_n) = 1$

4 Boosting

4.1

$\arg\min_{\beta_t} \epsilon_t(e^{\beta_t} - e^{-\beta_t}) + e^{-\beta_t}$

then:

$\frac{\partial}{\partial \beta_t} L = \epsilon_t e^{\beta_t} + \epsilon_t e^{-\beta_t} - e^{-\beta_t}$

set it to 0, we can get:

$$\begin{aligned}
\epsilon_t e^{\beta_t} + \epsilon_t e^{-\beta_t} - e^{-\beta_t} &= 0 \\
\epsilon_t e^{\beta_t} &= e^{-\beta_t} - \epsilon_t e^{-\beta_t} \\
e^{\beta_t} &= e^{-\beta_t}(\tfrac{1-\epsilon_t}{\epsilon_t}) \\
e^{2\beta_t} &= (\tfrac{1-\epsilon_t}{\epsilon_t}) \\
e^{\beta_t} &= (\tfrac{1-\epsilon_t}{\epsilon_t})^{\frac{1}{2}} \\
\beta_t &= \tfrac{1}{2}\ln(\tfrac{1-\epsilon_t}{\epsilon_t})
\end{aligned}$$

Therefore, the optimal $\beta_t^* = \frac{1}{2}\ln(\frac{1-\epsilon_t}{\epsilon_t})$

4.2

since the error of a random classifier is $\frac{1}{2}$,

$\epsilon_t = \sum_{n:y_n \neq h_t(\boldsymbol{x}_n)} D_t(n) = \frac{1}{2}$.

$\sum_{n:y_n \neq h_t(\boldsymbol{x}_n)} D_{t+1}(n) = \frac{D_t(n)\exp(-y_n\beta_t^* h_t^*)}{\sum_{n'=1}^{N} D_t(n)\exp(-y_n\beta_t^* h_t^*)}$

since $\beta_t^* = \frac{1}{2}\ln(\frac{1-\epsilon_t}{\epsilon_t})$ and $\epsilon_t = \frac{1}{2}$, we can get:

$\beta_t^* = 0$ then:

$\sum_{n:y_n \neq h_t(\boldsymbol{x}_n)} D_{t+1}(n) = \frac{\sum_{n:y_n \neq h_t(\boldsymbol{x}_n)} D_t(n)}{\sum_{n'=1}^{N} D_t(n)}$

since error of a random classifier is $\frac{1}{2}$, $\epsilon_t = \frac{1}{2}$,

$\sum_{n:y_n \neq h_t(\boldsymbol{x}_n)} D_{t+1}(n) = \frac{1}{2}$