hw3

wangxinshi47

March 2023

To find the subdifferetial of SVM, we use the affine transformation rule:

$$
\begin{aligned}
f(\omega, b) &= \frac{1}{n}\sum_{i=1}^{n}(1 - y_i(<\omega, x_i> + b))_+ + \frac{\lambda}{2}||\omega||_2^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(1 - y_i(\omega^T x_i + b))_+ + \frac{\lambda}{2}||\omega||_2^2 \\
\text{Let } h_i(\omega) &= (1 - y_i\omega^T x_i - y_i b)_+ \\
\partial h_i(\omega) &= -y_i x_i \cdot \partial p(z)|_{z=1-y_i x_i^T\omega - y_i b} \\
&= -y_i x_i \begin{cases} 0, & \text{if } 1 - y_i\omega^T x_i < y_i b \\ 1, & \text{if } 1 - y_i\omega^T x_i > y_i b \\ , & \text{if } 1 - y_i\omega^T x_i = y_i b \end{cases} \\
\partial f(\omega) &= \frac{1}{n}\sum_{i=1}^{n} -y_i x_i \begin{cases} 0, & \text{if } 1 - y_i\omega^T x_i < y_i b \\ 1, & \text{if } 1 - y_i\omega^T x_i > y_i b \\ , & \text{if } 1 - y_i\omega^T x_i = y_i b \end{cases} + \lambda\omega \\
\partial h_i(b) &= -y_i \cdot \partial p(z)|_{z=1-y_i x_i^T\omega - y_i b} \\
&= -y_i \begin{cases} 0, & \text{if } 1 - y_i\omega^T x_i < y_i b \\ 1, & \text{if } 1 - y_i\omega^T x_i > y_i b \\ , & \text{if } 1 - y_i\omega^T x_i = y_i b \end{cases} \\
\partial f(b) &= -y_i \begin{cases} 0, & \text{if } 1 - y_i\omega^T x_i < y_i b \\ 1, & \text{if } 1 - y_i\omega^T x_i > y_i b \\ , & \text{if } 1 - y_i\omega^T x_i = y_i b \end{cases}
\end{aligned}
$$

We can formulate this constrained optimization problem into the projected subgradient method with step size equal to α_t by removing constants and completing the squares.

$$\begin{aligned}
X_{t+1} =& \operatorname*{argmin}_{u \in C} f(X_t) + < g, u - X_t > + \frac{1}{2\alpha_t} ||u - X_t||_2^2 \\
& \text{We can remove } f(X_t) \text{ from the objective as it's a constant} \\
=& \operatorname*{argmin}_{u \in C} g^T(u - X_t) + \frac{1}{2\alpha_t}(u - X_t)^T(u - X_t) \\
& \text{We can multiply the objective by } 2\alpha_t \text{ as it's a constant} \\
=& \operatorname*{argmin}_{u \in C} 2\alpha_t g^T(u - X_t) + (u - X_t)^T(u - X_t) \\
& \text{We can add } \alpha_t^2 g^T g \text{ to the objective as it's a constant} \\
=& \operatorname*{argmin}_{u \in C} \alpha_t g^T(u - X_t) + \alpha_t g^T(u - X_t) + (u - X_t)^T(u - X_t) + \alpha_t^2 g^T g \\
=& \operatorname*{argmin}_{u \in C} \alpha_t g^T(u - X_t) + (\alpha_t g^T(u - X_t) + (u - X_t)^T(u - X_t)) + \alpha_t^2 g^T g \\
=& \operatorname*{argmin}_{u \in C} \alpha_t g^T(u - X_t) + (u - X_t + \alpha_t g)^T(u - X_t) + \alpha_t^2 g^T g \\
=& \operatorname*{argmin}_{u \in C} \alpha_t g^T(u - X_t + \alpha_t g) + (u - X_t + \alpha_t g)^T(u - X_t) \\
=& \operatorname*{argmin}_{u \in C} (u - X_t + \alpha_t g)^T(u - X_t + \alpha_t g) \\
=& \operatorname*{argmin}_{u \in C} ||(u - X_t + \alpha_t g)||_2^2 \\
=& P_c(X_t - \alpha_t g)
\end{aligned}$$

Therefore the constrained optimization problem is exactly the projected subgradient method.

To find the closed form solution for X_{t+1}, we set the derivative of the unconstrained problem to 0.

$$\begin{aligned}
h(P_t, X_t, g) &= \underset{u}{argmin} f(X_t) + < g, u - X_t > + \frac{1}{2\alpha_t} ||P_t^{1/2}(u - X_t)||_2^2 \\
\frac{\partial h}{\partial u} &= g + \frac{P_t^{1/2}}{\alpha_t}(P_t^{1/2}(u - X_t))
\end{aligned}$$

We set $\frac{\partial h}{\partial u} = 0$ to find the update scheme.

$$\begin{aligned}
g + \frac{P_t^{1/2}}{\alpha_t}(P_t^{1/2}(u - X_t)) &= 0 \\
\alpha_t g + P_t(u - X_t) &= 0 \\
P_t(u - X_t) &= -\alpha_t g \\
u - X_t &= -\alpha_t P_t^{-1} g \\
u &= X_t - \alpha_t P_t^{-1} g
\end{aligned}$$

Therefore $X_{t+1} = X_t - \alpha_t P_t^{-1} g$