CS6230: Optimization Methods in Machine Learning

HW2

Yash Chandarana
CS14BTECH11040

$$\underset{\beta\in\mathbb{R}^p,\beta_0\in\mathbb{R}\xi\in\mathbb{R}^n}{minimize}\frac{1}{2}||\beta||_2^2 + C_1\sum_{i\in S_1}\xi_i + C_2\sum_{i\in S_2}\xi_i$$

subject to

$$\xi_i \geq 0, y_i(x_i^T\beta+\beta_0) \geq 1-\xi_i \;\forall i = 1, 2, ..., n$$

where

$$S_1 = \{i \in \{1, ..., n\} : y_i = +1\}$$

and

$$S_2 = \{i \in \{1, ..., n\} : y_i = -1\}$$

1.
   - The function to optimize over is a sum of convex functions in the individual variables β, β_0 and ξ. Hence, the function is convex
   - $\beta = \mathbf{0}$, $\beta_0 = 0$ and $\xi = \mathbf{1}$ is a point of strict feasibility.
   - All the given constraints are affine in β, β_0 and ξ
     $\implies$ Slater's conditions are satisfied
     $\implies$ Strong duality holds

2. Let the lagrangian be

$$\mathcal{L}(\beta,\beta_0,\xi,\alpha,\mu) = \frac{1}{2}||\beta||_2^2+C_1\sum_{i\in S_1}\xi_i+C_2\sum_{i\in S_2}\xi_i+\sum_{i=1}^n\alpha_i(1-\xi_i-y_i(x_i^T\beta+\beta_0))-\sum_{i=1}^n\mu_i(\xi_i)$$

   - Using stationarity in the above, assuming $\beta^*, \beta_0^*, \xi^*$ to be the primal solutions and α^*, μ^* to be the dual solutions, we get:

$$\nabla_\beta\mathcal{L} = \beta + \sum_{i=1}^n\alpha_i(-y_ix_i) = \mathbf{0} \implies \beta = \sum_{i=1}^n\alpha_iy_ix_i \tag{1}$$

$$\frac{\partial\mathcal{L}}{\partial\beta_0} = \sum_{i=1}^n -\alpha_iy_i = 0 \implies \sum_{i=1}^n\alpha_iy_i = 0 \tag{2}$$

$$\frac{\partial \mathcal{L}}{\partial \xi_i} = C_1 I_{i \in S_1} + C_2 I_{i \in S_2} - \alpha_i - \mu_i = 0 \implies \alpha_i + \mu_i = C_j \; \forall i = 1, ..., n \; \forall j = 1, 2 \tag{3}$$

- Using complementary slackness

$$\alpha_i(1 - \xi_i - y_i(x_i^T \beta + \beta_0)) = 0$$

$$\mu_i \xi_i = 0$$

- Primal feasibility

$$\xi_i, y_i(x_i^T \beta + \beta_0) \geq 0$$

- Dual feasibility

$$\alpha_i, \mu_i \geq 0 \tag{4}$$

3. $\widetilde{X} = diag(y)X$
   From equation 1,

$$\beta = \widetilde{X}^T \alpha \implies ||\beta||_2^2 = \alpha^T \widetilde{X} \widetilde{X}^T \alpha \tag{5}$$

   Now, using equations 3 and 4, we can say that
   $0 \leq \alpha_i \leq C_1 I_{S_1} + C_2 I_{S_2} \implies \mathbf{0} \leq \alpha_{S_j} \leq C_j \mathbf{1} \forall j = 1, 2$
   Now, substitute equations 1, 2, 3 and 5 into equation for lagrangian to get,

$$\begin{aligned}
\mathcal{L} &= \frac{1}{2} \alpha^T \widetilde{X} \widetilde{X}^T \alpha + \sum_{i \in S_1} (C_1 - \alpha_i - \mu_i) \xi_i + \sum_{i \in S_2} (C_2 - \alpha_i - \mu_i) \xi_i \\
&\quad + \sum_{i=1}^{n} \alpha_i - \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i y_i \alpha_j y_j x_i^T x_j - \beta_0 \sum_{i=1}^{n} \alpha_i y_i \\
&= \frac{1}{2} \alpha^T \widetilde{X} \widetilde{X}^T \alpha + \sum_{i=1}^{n} \alpha_i - ||\beta||_2^2 \\
&= -\frac{1}{2} \alpha^T \widetilde{X} \widetilde{X}^T \alpha + \alpha^T \mathbf{1}
\end{aligned}$$

   So, we can write the dual as

$$\underset{\alpha \in \mathbb{R}^n}{maximise} - \frac{1}{2} \alpha^T \widetilde{X} \widetilde{X}^T \alpha + \alpha^T \mathbf{1}$$

   subject to $y^T \alpha = 0$, $0 \leq \alpha_{S_1} \leq C_1 \mathbf{1}$, $0 \leq \alpha_{S_2} \leq C_2 \mathbf{1}$

4. Both the above problems are quadratic problems with linear constraints