Department of Computer Science

Ramakrishna Mission Vivekananda Educational Research Institute, Belur Math

Optimization for Machine Learning, BDA 2023 batch

Problem set on Lagrangian Dual & KKT

Chapter 6, *Nonlinear Programming* by Bazaraa, Sherali and Shetty

Problem 6.10 (a,b), 6.11 (a,b), 6.16, 6.17, 6.26

***Additional problems*:**

1. Consider the problem

$$\text{minimize} f(x) = x_1^2 + x_2^2$$

$$\text{subject to: } (x_1 - 1)^2 + (x_2 - 1)^2 \le 1,\ (x_1 - 1)^2 + (x_2 + 1)^2 \le 1,\ x \in \mathbb{R}^2$$

(a) Sketch the feasible set and level sets of the objective function. Find optimal point $x^\star$ and optimal value $f(x^*)$.

(b) Give the KKT conditions. Do there exist Lagrange multipliers λ_1 and λ_2 that prove that $x^\star$ is optimal?

(c) Derive and solve the Lagrange dual problem. Does strong duality hold?

2. Solve the problem minimize $f(x) = x_1^2 + x_2^2$, subject to: $x_1 + 2x_2 \ge 2$, $x_1, x_2 \ge 0$. Find the Dual variables and verify if duality gap is there

3. The hard-margin SVM is used when data is linearly separable, i.e., where, $n = 1, \dots, N$ labeled data points are separated into, say, two classes $y_n = \{-1, +1\}$ by the hyperplane $\mathbf{w}^T\mathbf{x} + b$. The associated primal optimization problem is convex QP and given as

$$\begin{array}{lll} \underset{b,\mathbf{w}}{\text{minimize:}} & \frac{1}{2}\mathbf{w}^T\mathbf{w} & \\ \text{subject to:} & y_n(\mathbf{w}^T\mathbf{x}_n + b) \ge 1 & n = 1,\dots,N \end{array}$$

(a) Write the above problem in standard QP format

$$\begin{array}{ll} \underset{\mathbf{u}\in\mathbb{R}^L}{\text{minimize:}} & \frac{1}{2}\mathbf{u}^T\mathbf{Q}\mathbf{u} + \mathbf{p}^T\mathbf{u} \\ \text{subject to:} & \mathbf{A}\mathbf{u} \ge \mathbf{c}. \end{array}$$

find what are Q, p, A and c. Is Q a psd matrix?

(b) Derive the dual problem and show it is also QP of the following form

$$\begin{array}{ll} \underset{\alpha\in\mathbb{R}^N}{\text{minimize:}} & \frac{1}{2}\alpha^T \mathbf{Q}_D \alpha - \mathbf{1}_N^T \alpha \\ \text{subject to:} & \mathbf{A}_D \alpha \ge \mathbf{0}_{N+2}, \end{array}$$

find what are Q_D, A (D stands for dual). Is Q_D a psd matrix?

(c) Suppose instead of using the raw data we use a kernel function to transform it , i.e, we replace $\mathbf{x}$ by $\mathbf{z} = \Phi(\mathbf{x})$. What is the change in dual formulation, does Q_D still remain psd?

(d) Verify that for hard margin SVM, the min-max inequality holds with equality

(e) Derive the dual of the dual and show that it is same as the primal (Hint : consider the dual as primal and construct the Lagrangian, then derive its dual).

4. The most common formulation of the (linear) soft-margin SVM is allowing a margin of ξ_n for each data point $(\mathbf{x}_n, y_n)$ which captures by how much it fails to be separated

$$y_n\left(\mathrm{w}^{\mathrm{T}}\mathrm{x}_n + b\right) \geq 1 - \xi_n.$$

So, ξ_n captures how far into the margin the data point can go. Ideally, we would like the total sum of margin violations to be small, so we modify the hard-margin SVM to the soft-margin SVM by allowing margin violations but adding a penalty term to discourage large violations. The result is the soft-margin optimization problem:

$$\begin{array}{ll} \min\limits_{\mathbf{w},b,\boldsymbol{\xi}} & \frac{1}{2}\mathrm{w}^{\mathrm{T}}\mathrm{w} + C\sum\limits_{n=1}^{N}\xi_n \\ \text{subject to} & y_n\left(\mathrm{w}^{\mathrm{T}}\mathrm{x}_n + b\right) \geq 1 - \xi_n \text{ for } n = 1, 2, \ldots, N; \\ & \xi_n \geq 0 \text{ for } n = 1, 2, \ldots, N. \end{array}$$

a) Write the primal problem in standard QP format and show

$$\mathrm{Q} = \begin{bmatrix} 0 & \mathbf{0}_d^{\mathrm{T}} & \mathbf{0}_N^{\mathrm{T}} \\ \mathbf{0}_d & \mathrm{I}_d & \mathbf{0}_{d\times N} \\ \mathbf{0}_N & \mathbf{0}_{N\times d} & \mathbf{0}_{N\times N} \end{bmatrix},\ \mathbf{p} = \begin{bmatrix} \mathbf{0}_{d+1} \\ C\cdot\mathbf{1}_N \end{bmatrix},\ \mathrm{A} = \begin{bmatrix} \mathrm{YX} & \mathrm{I}_N \\ \mathbf{0}_{N\times(d+1)} & \mathrm{I}_N \end{bmatrix},\ \mathbf{c} = \begin{bmatrix} \mathbf{1}_N \\ \mathbf{0}_N \end{bmatrix}$$

b) Write the dual problem and verify that it is also a QP
c) Show that the only change of the dual problem is that each dual variable α_n is now upper-bounded by C, the penalty rate, instead of ∞.
d) What happens to optimal solution of soft-margin optimal hyperplane as the penalty parameter $C \rightarrow \infty$
e) For soft-margin SVM, show that the optimal b^* is not a fixed value, but it has a range of values which can be found from the optimal dual variables (Hint: for each n, we have $0 < \alpha^*_n < C$, using KKT conditions, find what happens to ξ_n when $\alpha^*_n > 0$ and $\alpha^*_n < C$)