ECE M146 | Discussion 4 Solution
Introduction to Machine Learning | Friday, April 26, 2019
Instructor: Lara Dolecek
TA: Zehui (Alex) Chen, Ruiyi (John) Wu

1. Introduction to optimization problem

   (a) Convex sets.

   (b) Convex functions.

   (c) Optimization problem in standard form.

   - Convex optimization.

   (d) Globally and locally optimal.

   (e) Duality.

   - Lagrange dual problem.
   - Geometric interpretation.
   - KKT conditions.

   Notes: One can read the book Convex Optimization by Boyd and Vandenberghe (freely available on-line) for more extensive coverage of the above topics.

2. Find the dual problem of the following Quadratic program

$$\begin{array}{ll} \text{minimize}_x & x^T P x \\ \text{subject to} & Ax \leq b \end{array}$$

Assume $P \in \mathcal{S}^n_{++}$.
Solution: The Lagrangian:

$$L(x, \lambda) = x^T P x + \lambda^T (Ax - b).$$

The dual function:

$$g(\lambda) = \text{inf}_x L(x, \lambda) = -\frac{1}{4}\lambda^T A P^{-1} A^T \lambda - b^T \lambda.$$

The dual problem

$$\begin{array}{ll} \text{maximize}_\lambda & -\dfrac{1}{4}\lambda^T A P^{-1} A^T \lambda - b^T \lambda \\ \text{subject to} & \lambda \geq 0. \end{array}$$

3. Quadratic program example Consider the objective function

$$J(x_1, x_2) = 5x_1^2 + 4x_1x_2 + 2x_2^2 + 2x_1 - 4x_2.$$

Find the optimal x that minimize $J(x)$ under the following constrains:

(a) No constrain.
Solution: Set the gradient of $J(x)$ to 0, we get:

$$\begin{aligned} 10x_1 + 4x_2 + 2 &= 0, \\ 4x_1 + 4x_2 - 4 &= 0. \end{aligned}$$

The above equation give use $x = [-1, 2]^T$.

(b) $x_1 + x_2 + 2 = 0$.
Solution: We use a Lagrange multiplier λ to enforce this constrain. $L(x, \lambda) = 5x_1^2 + 4x_1x_2 + 2x_2^2 + 2x_1 - 4x_2 + \lambda(x_1 + x_2 + 2)$. Setting the gradient with respect to x equals 0 and use this constrain, we get:

$$\begin{aligned} 10x_1 + 4x_2 + 2 + \lambda &= 0, \\ 4x_1 + 4x_2 - 4 + \lambda &= 0. \\ x_1 + x_2 + 2 &= 0. \end{aligned}$$

Solving the above equations, we get $x = [-1; -1]$ and $\lambda = 12$.

(c) $x_1 + x_2 + 2 \leq 0$.
Solution: Same as (b). See visualization.

(d) $x_1 + x_2 + 2 \geq 0$.
Solution: Same as (a). See visualization.

Visualization: See ppt for the visualization concerning this problem.

4. Relationship between soft SVM loss and logistic regression

   We learned that the loss function for the primal problem of SVM is of the form

$$\frac{1}{2}\|w\|^2 + C\sum_{n=1}^{m}\xi_n.$$

   We have seen that for data points that are on the correct side of the margin boundary, and which therefore satisfy $y_n h(x_n) \geq 1$, we have $\xi_n = 0$. For the remaining points we have $\xi_n = 1 - y_n h(x_n)$. Thus the objective function can be written in the form

$$\sum_{n=1}^{N} E_{SV}(y_n h(x_n)) + \lambda\|w\|^2.$$

   where $\lambda = (2C)^{-1}$, find out what is the error function $E_{SV}(.)$ in terms of $y_n h(x_n)$. **Solution:**

$$E_{SV}(y_n h(x_n)) = [1 - y_n h(x_n)]_+$$

   where [.] denotes the positive part. This error function is called the hinge error function.

   Compare this with the logistic regression loss function:

$$E_{LR}(y_n h(x_n)) = \ln(1 + \exp(-y_n h(x_n))).$$