EECS 553 HW 5

Due Thursday, October 3, by 11:59 PM via Gradescope and Canvas

When you upload your solutions to Gradescope, please indicate which pages of your solution are associated with each problem.

1. Constrained Optimization

In this exercise you will solve the following optimization problems using the KKT conditions.

$$\begin{aligned} \underset{\boldsymbol{x}\in\mathbb{R}^2}{\text{minimize}} \quad & (2x_1-1)^2+(x_2-2)^2 \\ \text{subject to} \quad & 3x_1+2x_2 \le 4 \\ & x_2 \ge x_1 \end{aligned}$$

(a) Explain why strong duality holds for this problem.

(b) Write down the Lagrangian and KKT conditions.

(c) By (a), to solve the primal problem it suffices to find $(x_1, x_2, \lambda_1, \lambda_2)$ satisfying the KKT conditions. Solve the primal problem by finding such values of $x_1, x_2, \lambda_1, \lambda_2$. *Hint:* For each inequality constraint, there are two cases: either the constraint is satisfied with equality, or the corresponding Lagrange multiplier must be zero. Since there are two inequality constraints, consider all four cases.

(d) Determine the Lagrangian dual function and write down the dual optimization problem. Then solve the original problem by first solving the dual problem and then inferring the primal solution from the dual solution.

2. Support Vector Regression

Support vector regression (SVR) is a method for regression analogous to the support vector classifier. Let $(\boldsymbol{x}_i, y_i) \in \mathbb{R}^d \times \mathbb{R}$, $i = 1, \dots, n$ be training data for a regression problem.

In the case of *linear regression*, SVR solves

$$\begin{aligned} \min_{\boldsymbol{w},b,\boldsymbol{\xi}^+,\boldsymbol{\xi}^-} \quad & \frac{1}{2}\|\boldsymbol{w}\|_2^2 + \frac{C}{n}\sum_{i=1}^{n}(\xi_i^+ + \xi_i^-) \\ \text{s.t.} \quad & y_i - \boldsymbol{w}^T\boldsymbol{x}_i - b \le \epsilon + \xi_i^+ \quad \forall i \\ & \boldsymbol{w}^T\boldsymbol{x}_i + b - y_i \le \epsilon + \xi_i^- \quad \forall i \\ & \xi_i^+ \ge 0 \quad \forall i \\ & \xi_i^- \ge 0 \quad \forall i \end{aligned}$$

where $\boldsymbol{w} \in \mathbb{R}^d$, $b \in \mathbb{R}$, $\boldsymbol{\xi}^+ = (\xi_i^+, \dots, \xi_n^+)^T$, and $\boldsymbol{\xi}^- = (\xi_i^-, \dots, \xi_n^-)^T$.

Here $C > 0$, $\epsilon > 0$ are fixed, and $\|\cdot\|_2$ is the Euclidean norm.

a. Show that for an appropriate choice of λ, SVR solves

$$\min_{\boldsymbol{w},b} \quad \frac{1}{n}\sum_{i=1}^{n} \ell_\epsilon(y_i, \boldsymbol{w}^T\boldsymbol{x}_i + b) + \lambda\|\boldsymbol{w}\|_2^2$$

where $\ell_\epsilon(y,t) = \max\{0, |y-t|-\epsilon\}$ is the so-called ϵ-*insensitive loss*, which does not penalize prediction errors below a level of ϵ.

Note: This part does not play a role in the subsequent parts.

b. The optimization problem is convex with affine constraints, and therefore strong duality holds. Use the KKT conditions to derive the dual optimization problem in a manner analogous to the support vector classifier. As in the SVC, you should eliminate the dual variables corresponding to the constraints $\xi_i^+ \geq 0$, $\xi_i^- \geq 0$.

c. Explain how to kernelize SVR. Be sure to explain how to determine b^* and evaluate the final prediction function.

d. Argue that the final prediction will only depend on a subset of training examples, and characterize those training examples. Your characterization should be analogous to the characterization of support vectors being "hard to classify" in support vector classification

3. Removing a Non-Support Vector (3 points each)

Suppose we apply a binary SVM classifier with kernel k to a dataset $(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)$. Suppose $\boldsymbol{\alpha}^*$ is a solution to the dual optimization problem, and that $\boldsymbol{x}_i$ is not a support vector for a certain i. This implies $\alpha_i^* = 0$.

Now consider the modified data set obtained by removing $(\boldsymbol{x}_i, y_i)$. Let $\boldsymbol{\alpha}_{-i}^*$ be the length $n-1$ vector obtained by removing the i-th entry from $\boldsymbol{\alpha}^*$.

(a) Show that $\boldsymbol{\alpha}_{-i}^*$ solves the dual for the modified dataset.

(b) Use part (a) to show that the SVM remains unchanged.