EECS 553 Homework 5 Solution (FA24)

1. Constrained Optimization

(a) (3 points)

Grading Rubrics

1. 1 point for stating that the objective function is convex.
2. 1 point for stating that the constraints are convex.
3. 1 point for verifying Slater's condition.

Solution:

Let the objective function be

$$f(\mathbf{x}) = (2x_1 - 1)^2 + (x_2 - 2)^2.$$

The Hessian of $f(\mathbf{x})$ is

$$\nabla^2 f(\mathbf{x}) = \begin{bmatrix} 8 & 0 \\ 0 & 2 \end{bmatrix},$$

which is positive definite since its eigenvalues are 8 and 2. Therefore, $f(\mathbf{x})$ is convex.

The constraints are:

$$g_1(\mathbf{x}) = 3x_1 + 2x_2 - 4 \leq 0,$$

$$g_2(\mathbf{x}) = x_1 - x_2 \leq 0.$$

Both g_1 and g_2 are affine functions, hence convex.

To verify Slater's condition, we can pick $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Thus, Slater's condition is satisfied, and strong duality holds for this problem.

(b) (3 points)

Grading Rubrics

1. 1 point for writing the correct Lagrangian.
2. 2 points for writing the KKT conditions.

Solution:

Introduce Lagrange multipliers $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$ corresponding to g_1 and g_2, respectively. The Lagrangian is

$$L(\mathbf{x}, \boldsymbol{\lambda}) = (2x_1 - 1)^2 + (x_2 - 2)^2 + \lambda_1(3x_1 + 2x_2 - 4) + \lambda_2(x_1 - x_2).$$

KKT Conditions:

Stationarity:

$$\frac{\partial L}{\partial x_1} = 4(2x_1 - 1) + 3\lambda_1 + \lambda_2 = 0, \quad (1)$$
$$\frac{\partial L}{\partial x_2} = 2(x_2 - 2) + 2\lambda_1 - \lambda_2 = 0. \quad (2)$$

Primal Feasibility:

$$3x_1 + 2x_2 - 4 \leq 0, \quad (3)$$
$$x_1 - x_2 \leq 0. \quad (4)$$

Dual Feasibility:

$$\lambda_1 \geq 0, \quad \lambda_2 \geq 0. \quad (5)$$

Complementary Slackness:

$$\lambda_1(3x_1 + 2x_2 - 4) = 0, \quad (6)$$
$$\lambda_2(x_1 - x_2) = 0. \quad (7)$$

(c) (3 points)

Grading Rubrics

1. 2 points for identifying the four cases.
2. 1 point for identifying the optimal solution among the cases.

Solution:

We consider all four combinations of the constraints being active (equality holds) or inactive (Lagrange multiplier is zero).

Case 1: $\lambda_1 = 0$, $\lambda_2 = 0$

From (1) and (2):

$$4(2x_1 - 1) = 0 \implies x_1 = \tfrac{1}{2},$$
$$2(x_2 - 2) = 0 \implies x_2 = 2.$$

Check primal feasibility:

$$3\left(\tfrac{1}{2}\right) + 2(2) - 4 = \tfrac{3}{2} + 4 - 4 = \tfrac{3}{2} > 0,$$

which violates constraint (3). Thus, this case is infeasible.

Case 2: $\lambda_1 = 0$, $\lambda_2 > 0$

From complementary slackness (7):

$$x_1 - x_2 = 0 \implies x_1 = x_2.$$

From (1) and (2):

$$4(2x_1 - 1) + \lambda_2 = 0,$$
$$2(x_1 - 2) - \lambda_2 = 0.$$

Adding the equations:

$$4(2x_1 - 1) + 2(x_1 - 2) = 0 \implies 10x_1 - 8 = 0 \implies x_1 = \tfrac{4}{5}.$$

Then $\lambda_2 = 2(x_1 - 2) = 2\left(\tfrac{4}{5} - 2\right) = -\tfrac{12}{5} < 0$, which violates dual feasibility (5). Thus, this case is invalid.

Case 3: $\lambda_1 > 0$, $\lambda_2 = 0$

From complementary slackness (6):

$$3x_1 + 2x_2 - 4 = 0. \quad (8)$$

From (1) and (2):

$$\begin{aligned} 8x_1 - 4 + 3\lambda_1 &= 0, \quad (9)\\ 2x_2 - 4 + 2\lambda_1 &= 0. \quad (10) \end{aligned}$$

Solve (9) and (10) for x_1 and x_2:

$$\begin{aligned} x_1 &= \tfrac{4-3\lambda_1}{8},\\ x_2 &= 2 - \lambda_1. \end{aligned}$$

Substitute into (8):

$$3\left(\tfrac{4-3\lambda_1}{8}\right) + 2(2 - \lambda_1) - 4 = 0 \implies \lambda_1 = \tfrac{12}{25}.$$

Compute x_1 and x_2:

$$x_1 = \frac{4-3(\frac{12}{25})}{8} = \tfrac{8}{25}, \quad x_2 = 2 - \tfrac{12}{25} = \tfrac{38}{25}.$$

Check primal feasibility:

$$x_1 - x_2 = \tfrac{8}{25} - \tfrac{38}{25} = -\tfrac{30}{25} = -\tfrac{6}{5} \leq 0.$$

Dual feasibility is satisfied since $\lambda_1 = \tfrac{12}{25} > 0$ and $\lambda_2 = 0 \geq 0$. This case provides a feasible solution.

Case 4: $\lambda_1 > 0$, $\lambda_2 > 0$

From complementary slackness:

$$x_1 - x_2 = 0 \implies x_1 = x_2, \quad 3x_1 + 2x_2 - 4 = 0 \implies 5x_1 - 4 = 0 \implies x_1 = \tfrac{4}{5}.$$

From (1) and (2):

$$\begin{aligned} 8x_1 - 4 + 3\lambda_1 + \lambda_2 &= 0,\\ 2x_1 - 4 + 2\lambda_1 - \lambda_2 &= 0. \end{aligned}$$

Subtracting the equations:

$$(8x_1 - 4) - (2x_1 - 4) + (3\lambda_1 + \lambda_2) - (2\lambda_1 - \lambda_2) = 0 \implies 6x_1 + \lambda_1 + 2\lambda_2 = 0.$$

Substituting $x_1 = \tfrac{4}{5}$:

$$6\left(\tfrac{4}{5}\right) + \lambda_1 + 2\lambda_2 = 0 \implies \lambda_1 + 2\lambda_2 = -\tfrac{24}{5} < 0,$$

which violates dual feasibility (5). Thus, this case is invalid.

Conclusion: The optimal solution is:

$$x_1^* = \tfrac{8}{25}, \quad x_2^* = \tfrac{38}{25}, \quad \lambda_1^* = \tfrac{12}{25}, \quad \lambda_2^* = 0.$$

(d) (3 points)

Grading Rubrics

1. 1 point for writing the dual problem.
2. 1 point for solving the dual problem.
3. 1 point for inferring the primal solution from the dual solution.

Solution:

The Lagrangian is:

$$L(\mathbf{x}, \boldsymbol{\lambda}) = (2x_1 - 1)^2 + (x_2 - 2)^2 + \lambda_1(3x_1 + 2x_2 - 4) + \lambda_2(x_1 - x_2).$$

To find the dual function $g(\boldsymbol{\lambda})$, minimize L over $\mathbf{x} \in \mathbb{R}^2$ for given $\boldsymbol{\lambda} \geq 0$.

Stationarity Conditions:

$$\frac{\partial L}{\partial x_1} = 8x_1 - 4 + 3\lambda_1 + \lambda_2 = 0,$$
$$\frac{\partial L}{\partial x_2} = 2x_2 - 4 + 2\lambda_1 - \lambda_2 = 0.$$

Solving for x_1 and x_2:

$$x_1 = \tfrac{4 - 3\lambda_1 - \lambda_2}{8},$$
$$x_2 = 2 - \lambda_1 + \tfrac{\lambda_2}{2}.$$

Substitute x_1 and x_2 back into L to obtain $g(\boldsymbol{\lambda})$:

$$g(\boldsymbol{\lambda}) = -\tfrac{(25\lambda_1^2 - 10\lambda_1\lambda_2 + 5\lambda_2^2)}{16}.$$

Dual Problem:

$$\underset{\lambda_1, \lambda_2 \geq 0}{\text{maximize}} \quad g(\lambda_1, \lambda_2) = -\tfrac{25\lambda_1^2 - 10\lambda_1\lambda_2 + 5\lambda_2^2}{16}.$$

To maximize $g(\boldsymbol{\lambda})$, we set the gradient to zero:

$$\frac{\partial g}{\partial \lambda_1} = -\tfrac{1}{16}(50\lambda_1 - 10\lambda_2) = 0,$$
$$\frac{\partial g}{\partial \lambda_2} = -\tfrac{1}{16}(-10\lambda_1 + 10\lambda_2) = 0.$$

Solving:

$$50\lambda_1 - 10\lambda_2 = 0, \quad -10\lambda_1 + 10\lambda_2 = 0 \implies \lambda_1 = \lambda_2 = 0.$$

So the maximizer lies on the boundaries. By testing $\lambda_1 = 0, \lambda_2 \neq 0$ and $\lambda_1 \neq 0, \lambda_2 = 0$, we can conclude that the maximum occurs at $\lambda_1 = \frac{12}{25}$, $\lambda_2 = 0$, consistent with the primal solution. By solving the dual problem, we confirm the primal optimal solution:

$$x_1^* = \tfrac{8}{25}, \quad x_2^* = \tfrac{38}{25}.$$

2. Support Vector Regression

a. (3 pts)

Grading Rubrics

(1) 1 points each for claiming equality formula for ξ_i^+ and ξ_i^- (i.e., $\xi_i^+ = \max\{0, |y_i - t_i| - \epsilon\}$ and $\xi_i^- = \max\{0, |y_i - t_i| - \epsilon\}$)

(2) 1 point for showing $\lambda = 1/2C$ is the appropriate choice

Solution:

The objective function in SVR is

$$\frac{1}{2}\|\mathbf{w}\|_2^2 + \frac{C}{n}\sum_{i=1}^{n}(\xi_i^+ + \xi_i^-).$$

Subject to the constraints:

$$\begin{aligned} y_i - \mathbf{w}^\top \mathbf{x}_i - b &\leq \epsilon + \xi_i^+, \quad \forall i, \\ \mathbf{w}^\top \mathbf{x}_i + b - y_i &\leq \epsilon + \xi_i^-, \quad \forall i, \\ \xi_i^+ &\geq 0, \quad \forall i, \\ \xi_i^- &\geq 0, \quad \forall i. \end{aligned}$$

From the first constraint, we have

$$\xi_i^+ \geq y_i - \mathbf{w}^\top \mathbf{x}_i - b - \epsilon.$$

Since $\xi_i^+ \geq 0$, this implies

$$\xi_i^+ = \max\{0, y_i - (\mathbf{w}^\top \mathbf{x}_i + b) - \epsilon\}.$$

Similarly, from the second constraint,

$$\xi_i^- = \max\{0, (\mathbf{w}^\top \mathbf{x}_i + b) - y_i - \epsilon\}.$$

Therefore, the sum of the slack variables can be written as

$$\xi_i^+ + \xi_i^- = \max\{0, |y_i - (\mathbf{w}^\top \mathbf{x}_i + b)| - \epsilon\} = \ell_\epsilon(y_i, \mathbf{w}^\top \mathbf{x}_i + b).$$

Substituting back into the objective function, we get

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|_2^2 + \frac{C}{n}\sum_{i=1}^{n}\ell_\epsilon(y_i, \mathbf{w}^\top \mathbf{x}_i + b).$$

This is equivalent to

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

where $\lambda = \frac{1}{2C}$. Thus, SVR solves the given optimization problem with the appropriate choice of λ.

b. (3 pts)

Grading Rubrics

(1) 1 point for Lagrangian

(2) 0.25 points each for 4 stationary condition equations (total 1 point)

(3) 1 point for the final dual problem

Solution:
We begin by forming the Lagrangian L:

$$\begin{aligned} L(\mathbf{w}, b, \boldsymbol{\xi}^+, \boldsymbol{\xi}^-, \boldsymbol{\alpha}^+, \boldsymbol{\alpha}^-, \boldsymbol{\beta}^+, \boldsymbol{\beta}^-) = &\frac{1}{2}\|\mathbf{w}\|_2^2 + \frac{C}{n}\sum_{i=1}^n (\xi_i^+ + \xi_i^-) \\ &+ \sum_{i=1}^n \alpha_i^+[y_i - \mathbf{w}^\top \mathbf{x}_i - b - \epsilon - \xi_i^+] \\ &+ \sum_{i=1}^n \alpha_i^-[\mathbf{w}^\top \mathbf{x}_i + b - y_i - \epsilon - \xi_i^-] \\ &- \sum_{i=1}^n (\beta_i^+ \xi_i^+ + \beta_i^- \xi_i^-), \end{aligned}$$

where $\alpha_i^+, \alpha_i^- \geq 0$ are the Lagrange multipliers for the inequality constraints, and $\beta_i^+, \beta_i^- \geq 0$ for the non-negativity constraints on ξ_i^+, ξ_i^-.

Stationarity Conditions:

1. Gradient w.r.t. $\mathbf{w}$:

$$\frac{\partial L}{\partial \mathbf{w}} = \mathbf{w} - \sum_{i=1}^n \alpha_i^+(-\mathbf{x}_i) + \sum_{i=1}^n \alpha_i^- \mathbf{x}_i = 0 \implies \mathbf{w} = \sum_{i=1}^n (\alpha_i^+ - \alpha_i^-)\mathbf{x}_i.$$

2. Gradient w.r.t. b:

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^n \alpha_i^+ - \sum_{i=1}^n \alpha_i^- = 0 \implies \sum_{i=1}^n (\alpha_i^+ - \alpha_i^-) = 0.$$

3. Gradient w.r.t. ξ_i^+:

$$\frac{\partial L}{\partial \xi_i^+} = \frac{C}{n} - \alpha_i^+ - \beta_i^+ = 0.$$

4. Gradient w.r.t. ξ_i^-:

$$\frac{\partial L}{\partial \xi_i^-} = \frac{C}{n} - \alpha_i^- - \beta_i^- = 0.$$

Since $\beta_i^+, \beta_i^- \geq 0$, it follows that

$$0 \leq \alpha_i^+ \leq \frac{C}{n}, \quad 0 \leq \alpha_i^- \leq \frac{C}{n}.$$

Dual Function:
Substituting $\mathbf{w}$ into the Lagrangian, we obtain the dual function:

$$L_D(\boldsymbol{\alpha}^+, \boldsymbol{\alpha}^-) = -\frac{1}{2}\sum_{i,j=1}^n (\alpha_i^+ - \alpha_i^-)(\alpha_j^+ - \alpha_j^-)\mathbf{x}_i^\top \mathbf{x}_j + \sum_{i=1}^n (\alpha_i^+ - \alpha_i^-)y_i - \epsilon \sum_{i=1}^n (\alpha_i^+ + \alpha_i^-).$$

Dual Problem:

The dual optimization problem is:

$$\begin{aligned}
\max_{\boldsymbol{\alpha}^+,\boldsymbol{\alpha}^-} \quad & -\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i^+ - \alpha_i^-)(\alpha_j^+ - \alpha_j^-)\mathbf{x}_i^\top \mathbf{x}_j + \sum_{i=1}^{n}(\alpha_i^+ - \alpha_i^-)y_i - \epsilon\sum_{i=1}^{n}(\alpha_i^+ + \alpha_i^-), \\
\text{s.t.} \quad & \sum_{i=1}^{n}(\alpha_i^+ - \alpha_i^-) = 0, \\
& 0 \le \alpha_i^+, \alpha_i^- \le \frac{C}{n}, \quad \forall i.
\end{aligned}$$

This is a convex quadratic programming problem.

c. (3 pts)

Grading Rubrics

(1) 1 point for mentioning changing inner product of $\mathbf{x}_i, \mathbf{x}_j$ to $k(\mathbf{x}_i, \mathbf{x}_j)$

(2) 1 point for the two expressions of b^* depending on the dual variables

(3) 1 point for kernelizing the predictor

Solution:

To kernelize SVR, we replace the inner products $\mathbf{x}_i^\top \mathbf{x}_j$ with a kernel function $k(\mathbf{x}_i, \mathbf{x}_j)$. The dual problem becomes:

$$\begin{aligned}
\max_{\boldsymbol{\alpha}^+,\boldsymbol{\alpha}^-} \quad & -\frac{1}{2}\sum_{i,j=1}^{n}(\alpha_i^+ - \alpha_i^-)(\alpha_j^+ - \alpha_j^-)k(\mathbf{x}_i, \mathbf{x}_j) + \sum_{i=1}^{n}(\alpha_i^+ - \alpha_i^-)y_i - \epsilon\sum_{i=1}^{n}(\alpha_i^+ + \alpha_i^-), \\
\text{s.t.} \quad & \sum_{i=1}^{n}(\alpha_i^+ - \alpha_i^-) = 0, \\
& 0 \le \alpha_i^+, \alpha_i^- \le \frac{C}{n}, \quad \forall i.
\end{aligned}$$

The predictor function is kernelized as:

$$f(\mathbf{x}) = \sum_{i=1}^{n}(\alpha_i^+ - \alpha_i^-)k(\mathbf{x}_i, \mathbf{x}) + b^*.$$

To determine b^*, we utilize the complementary slackness. For any support vector with $0 < \alpha_i^+ < \frac{C}{n}$, we have:

$$y_i - f(\mathbf{x}_i) = \epsilon \implies b^* = y_i - \sum_{j=1}^{n}(\alpha_j^+ - \alpha_j^-)k(\mathbf{x}_j, \mathbf{x}_i) - \epsilon.$$

Similarly, for $0 < \alpha_i^- < \frac{C}{n}$:

$$f(\mathbf{x}_i) - y_i = \epsilon \implies b^* = y_i - \sum_{j=1}^{n}(\alpha_j^+ - \alpha_j^-)k(\mathbf{x}_j, \mathbf{x}_i) + \epsilon.$$

In practice, b^* is computed by averaging b^* over all such support vectors.

d. (3 pts)

Grading Rubrics

(1) 1.5 point each for two cases when the training samples are support vectors

The final predictor will only depend on those training samples for which either α_i^+ or α_i^- are nonzero. We call these *support vectors*, and they are those points for which either

$$y_i - w^T x - b = \epsilon + \xi_i^+$$

or

$$b - w^T x - y_i = \epsilon + \xi_i^-$$

hold, or in other words,

The training points with absolute prediction errors at least ϵ.

3. Removing a Non-Support Vector

a. (3 pts) **Grading Rubrics**

(1) 1 point to state the removal of (x_i, y_i) is equivalent to enforcing an extra constraint.

(2) 1 point to state that the optimized dual after the removal is no greater than the optimized objective of the dual.

(3) 1 point for the observation that $\boldsymbol{\alpha}^*$ makes them equal.

(4) Other correct proofs shall receive full credits.

The removal of (x_i, y_i) is equivalent to enforcing an extra constraint $\alpha_i = 0$ on the original problem. So the optimized objective of the dual after the removal is no greater than the optimized objective of the dual before the removal (because of the extra constraint). Yet $\boldsymbol{\alpha}^*$ makes them equal, so $\boldsymbol{\alpha}^*_{-i}$ solves the dual after the removal.

b. (3 pts) **Grading Rubrics**

(1) 1 point for writing solutions to the problems before and after the removal.

(1) 2 points for showing they are equal.

(4) Other correct proofs shall receive full credits.

$$\begin{aligned} f(\boldsymbol{x}) &= \text{sign}(\sum_{j=1}^{n} \alpha_j^* y_j k(\boldsymbol{x}_j, \boldsymbol{x}) + b^*) \\ &= \text{sign}(\sum_{j=1, j\neq i}^{n} \alpha_j^* y_j k(\boldsymbol{x}_j, \boldsymbol{x}) + b^*_{-i}) \end{aligned}$$

where $b^*_{-i} = y_k - \sum_{j=1,j\neq i}^{n} \alpha_j^* y_j k(\boldsymbol{x}_j, \boldsymbol{x}_k) = y_k - \sum_{j=1}^{n} \alpha_j^* y_j k(\boldsymbol{x}_j, \boldsymbol{x}_k) = b^*$ for k with $\alpha_k^* \in (0, \frac{C}{n})$. By (a), the first line is the solution to the original problem and the second line is the solution to the problem after the removal.