EECS 553 HW5

Lingqi Huang

October 2024

1 Problem 1

Part(a): Notice that the function of $f(x_1, x_2) = (2x_1 - 1)^2 + (x_2 - 2)^2$ is convex, and we notice the constraint function $g_1(x_1, x_2) = 3x_1 + 2x_2 - 4$ and $g_2(x_1, x_2) = x_1 - x_2$ are all confine, and thus the constraint qualification holds. By **Theorem 1** in Lecture note with title "Constraint Optimization":, we conclude that $p^* = d^*$, and thus the strong duality holds.

Part(b): The Lagrangian can be written as

$$L(x_1, x_2, \lambda_1, \lambda_2) = (2x_1 - 1)^2 + (x_2 - 2)^2 + \lambda_1(3x_1 + 2x_2 - 4) + \lambda_2(x_1 - x_2)$$

We first remind the KKT condition:
Thus, (1) implies

(1) $\nabla_{\boldsymbol{u}} f(\boldsymbol{u}^*) + \sum_{i=1}^{r} \lambda_i^* \nabla_{\boldsymbol{u}} g_i(\boldsymbol{u}^*) + \sum_{j=1}^{s} \nu_j^* \nabla_{\boldsymbol{u}} h_j(\boldsymbol{u}^*) = 0$

(2) $g_i(\boldsymbol{u}^*) \leq 0 \ \forall i$

(3) $h_j(\boldsymbol{u}^*) = 0 \ \forall j$

(4) $\lambda_i^* \geq 0 \ \forall i$

(5) $\lambda_i^* g_i(\boldsymbol{u}^*) = 0 \ \forall i$ *(complimentary slackness).*

$$\nabla_{\mathbf{x}} L = \begin{pmatrix} 4(2x_1 - 1) + 3\lambda_1 + \lambda_2 \\ 2(x_2 - 2) + 2\lambda_1 - \lambda_2 \end{pmatrix} = \begin{pmatrix} 8x_1 + 3\lambda_1 + \lambda_2 - 4 \\ 2x_2 + 2\lambda_1 - \lambda_2 - 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

(2) implies that $g_1(\mathbf{x}^*) \leq 0, g_2(\mathbf{x}^*) \leq 0$. Condition (3) is automatically satisfied. (4) implies $\lambda_1^* \geq 0, \lambda_2^* \geq 0$. (5) implies $\lambda_1^*(3x_1^* + 2x_2^* - 4) \leq 0, \lambda_2(x_1^* - x_2^*) \leq 0$.

Part(c): we need to consider all four cases:
Case 1:

$$\lambda_1^* = \lambda_2^* = 0$$

$$3x_1^* + 2x_2^* - 4 \leq 0$$

$$x_1^* - x_2^* \leq 0$$

Combining with the condition (1) of KKT, we will then get the solution of

$$\begin{cases} x_1^* = \frac{1}{2} \\ x_2^* = 2 \\ \lambda_1^* = 0 \\ \lambda_2^* = 0 \end{cases}$$

Notice this solution is invalid because this does not satisfy the constraint presented in (2) because $3x_1^* + 2x_2^* - 4 > 0$.

Case 2:

$$\lambda_1^* = 0$$

$$\lambda_2^* \geq 0$$

$$3x_1^* + 2x_2^* - 4 \leq 0$$

$$x_1^* - x_2^* = 0$$

Combining with condition (1) of KKT, we will solve that

$$\begin{cases} x_1^* = \frac{4}{5} \\ x_2^* = \frac{4}{5} \\ \lambda_1^* = 0 \\ \lambda_2^* = -\frac{12}{5} \end{cases}$$

Notice this solution is invalid because it violates the assumption of $\lambda_2^* \geq 0$.

Case 3:

$$\lambda_1^* \geq 0$$

$$\lambda_2 = 0$$

$$3x_1^* + 2x_2^* - 4 = 0$$

$$x_1^* - x_2^* \leq 0$$

Combining with condition (1) of KKT, we get the solution of

$$\begin{cases} x_1^* = \frac{8}{25} \\ x_2^* = \frac{38}{25} \\ \lambda_1^* = \frac{12}{25} \\ \lambda_2^* = 0 \end{cases}$$

This is a valid solution.

Case 4:

$$\lambda_1^* \geq 0$$

$$\lambda_2^* \geq 0$$

$$3x_1^* + 2x_2^* - 4 = 0$$

$$x_1^* - x_2^* = 0$$

Combining with condition (1) of KKT, we get the solution of

$$\begin{cases} x_1^* = \frac{4}{5} \\ x_2^* = \frac{4}{5} \\ \lambda_1^* = 0 \\ \lambda_2^* = \frac{12}{5} \end{cases}$$

We observe that both case3 and case4 give us the valid solution, however, case3 would give us the lowest objective value which is $\frac{9}{25}$ So the solution in case3 is what we desired.

Part(d): The dual function would be written as

$$L_D(x_1, x_2, \lambda_1, \lambda_2) = \min_{x_1, x_2} 2(x_1 - 1)^2 + (x_2 - 2)^2 + \lambda_1(3x_1 + 2x_2 - 4) + \lambda_2(x_1 - x_2)$$

By setting $\nabla_{\mathbf{x}} L_D = 0$, we would solve that

$$\begin{cases} x_1^* = \frac{4 - 3\lambda_1 - \lambda_2}{8} \\ x_2^* = 2 - \lambda_1 + \frac{\lambda_2}{2} \end{cases}$$

Now if we plug x_1^*, x_2^* into L_D and maximize it, we would need to solve the optimization problem of

$$\max_{\lambda_1 \geq 0, \lambda_2 \geq 0} -\frac{16}{25}\lambda_1^2 - \frac{5}{16}\lambda_2^2 + \frac{5}{8}\lambda_1\lambda_2 + \frac{3}{2}\lambda_1 - \frac{3}{2}\lambda_2$$

Now if we compute the gradient of the objective function and set it to 0, we would get that

$$\begin{cases} \lambda_1^* = 0 \\ \lambda_2^* = -\frac{12}{5} \end{cases}$$

But this did not satisfy with the constraint, and since we know there is only one critical point, we must achieve maximum at the boundary of the constraint, that either $\lambda_1 = \lambda_2 = 0$, or $\lambda_2 = 0$ and we need to solve for λ_1. Now if $\lambda_1 = \lambda_2 = 0$, we get the objective value of 0, and if we fix $\lambda_2 = 0$ and find the λ_1 that maximize the objective function, we could easily take derivative and get $\lambda_1 = \frac{12}{25}$, and we reached the objective value of $d^* = \frac{9}{25}$. Now by strong duality, we must have the $p^* = d^*$, and then if we plug derived $\lambda_1^* = \frac{12}{25}$ and $\lambda_2^* = 0$ back to get x_1^* and x_2^*, we will finally get the primal optimal is

$$\begin{cases} x_1^* = \frac{8}{25} \\ x_2^* = \frac{38}{25} \end{cases}$$

which is consistent with what we get from part(c).

2 Problem 2

Part(a): Notice that the constraint could be re-written as

$$\xi_i^+ \geq \max\{0, y_i - w^T x_i - b - \epsilon\}$$

$$\xi_i^- \geq \max\{0, w^T x_i + b - y_i - \epsilon\}$$

Thus the optimization problem can be re-written as

$$\min_{w,b} \frac{1}{2}||w||^2 + \frac{C}{n}\sum_{i=1}^{n} \max\{0, y_i - w^T x_i - b - \epsilon\} + \max\{0, w^T x_i + b - y_i - \epsilon\}$$

Notice that each term inside of summation can be re-written as $\max\{0, |y - w^T x_i - b| - \epsilon\}$, and if we pick $\lambda = \frac{1}{2C}$, we get the objective function can be written as

$$\min_{w,b} \frac{1}{n}\sum_{i=1}^{n} l_\epsilon(y, w^T x_i + b) + \lambda||w||^2$$

where $l_\epsilon = \max\{0, |y - t| - \epsilon\}$.

Part(b): We will first define the Lagrangian to be

$$\begin{aligned} L(w, b, \boldsymbol{\xi}^+, \boldsymbol{\xi}^-, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}) = \frac{1}{2}||w||^2 + \frac{C}{n}\sum_{i=1}^{n}(\xi_i^+ + \xi_i^-) + \sum_{i=1}^{n}\alpha_i(y_i - \langle w, x_i\rangle - b - \epsilon - \xi_i^+) \quad (1) \\ - \sum_{i=1}^{n}\beta_i(y_i - \langle w, x_i\rangle - b + \epsilon + \xi_i^-) - \sum_{i=1}^{n}\gamma_i\xi_i^+ - \sum_{i=1}^{n}\eta_i\xi_i^- \quad (2) \end{aligned}$$

where $\boldsymbol{\xi}^+, \boldsymbol{\xi}^-, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}$ are all vectors with length of n. Thus, the dual optimization problem would be

$$\max_{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}} \min_{w, b, \boldsymbol{\xi}^+, \boldsymbol{\xi}^-} L(w, b, \boldsymbol{\xi}^+, \boldsymbol{\xi}^-, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta})$$

We first fix $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}$ and first solve for the dual problem, that is we first take partial derivatives to all w, b, ξ_i^+, ξ_i^- and set then to 0, we will then get:

$$\begin{cases} w = \sum\limits_{i=1}^{n} (\alpha_i - \beta_i) x_i \\ \sum\limits_{i=1}^{n} \alpha_i = \sum\limits_{i=1}^{n} \beta_i \\ \alpha_i + \gamma_i = \frac{C}{n} \qquad \forall i \\ \beta_i + \eta_i = \frac{C}{n} \qquad \forall i \end{cases}$$

Notice that if one of the last two constraint does not hold we would reach $L_D(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}) = -\infty$.Now if we plug all four equality into L, we will see all $\xi^+ i, \xi^- i$ are canceled, and we will result in the dual optimization problem that is

$$\max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} -\frac{1}{2} \sum_{i,j=1}^{n} (\alpha_i - \beta_i)(\alpha_j - \beta_j) x_i^T x_j + \sum_{i=1}^{n} (\alpha_i - \beta_i) y_i - \sum_{i=1}^{n} (\alpha_i + \beta_i) \epsilon$$

$$st. \qquad \sum_{i=1}^{n} (\alpha_i - \beta_i) = 0 \tag{3}$$

$$0 \le \alpha_i \le \frac{C}{n} \tag{4}$$

$$0 \le \beta_i \le \frac{C}{n} \tag{5}$$

$$\tag{6}$$

Part(c): To kernalize SVR, we can simply replace $x_i^T x_j$ into $k(x_i, x_j)$, that is

$$\max_{\boldsymbol{\alpha}, \boldsymbol{\beta}} -\frac{1}{2} \sum_{i,j=1}^{n} (\alpha_i - \beta_i)(\alpha_j - \beta_j) k(x_i, x_j) + \sum_{i=1}^{n} (\alpha_i - \beta_i) y_i - \sum_{i=1}^{n} (\alpha_i + \beta_i) \epsilon$$

$$st. \qquad \sum_{i=1}^{n} (\alpha_i - \beta_i) = 0 \tag{7}$$

$$0 \le \alpha_i \le \frac{C}{n} \tag{8}$$

$$0 \le \beta_i \le \frac{C}{n} \tag{9}$$

$$\tag{10}$$

After solving for $\boldsymbol{\alpha}, \boldsymbol{\beta}$ by QP in SVR, in order to determine b^*, notice that by complimentary slackness, $\forall j$ such that $0 < \alpha_j < \frac{C}{n}$, we must have $\xi_j^+ = 0$ because again by complimentary slackness that $\gamma_j > 0$. So we must have for any j, such that $0 < \alpha_j < \frac{C}{n}$,

$$b^* = y_j - \sum_{i=1}^{n} (\alpha_i - \beta_i) \langle x_i, x_j \rangle - \epsilon$$

We could also do similar thing in term of ξ_i^+. The procedure of getting b^* of kernalized SVR is first solve b in single SVR term that by the step we described above, and then substitute the inner product with kernal function, that is, $\forall j$ such that $0 < \alpha_j < \frac{C}{n}$,

$$b^* = y_j - \sum_{i=1}^{n} (\alpha_i - \beta_i) k(x_i, x_j) - \epsilon$$

and then we substitute the inner product into kernal function, which is what we get above. Also, the final prediction function would be

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \beta_i) k(x_i, x_j) + b$$

Part(d):Notice that by complimentary slackness, for any i, we must have

$$\alpha_i(y_i - \langle w^*, x_i\rangle - b^* - \epsilon - \xi_i^+) = 0$$

$$\beta_i(y_i - \langle w^*, x_i\rangle - b^* + \epsilon + \xi_i^-) = 0$$

Thus, if x_i satisfy one of

$$y_i - \langle w^*, x_i\rangle - b^* - \epsilon = \xi_i^+$$

or

$$y_i - \langle w^*, x_i\rangle - b^* + \epsilon = -\xi_i^-$$

Then x_i must be a support vector. Since we know both ξ_i^+ and ξ_i^- are non-negative, then:
Case 1: If

$$y_i - \langle w^*, x_i\rangle - b^* + \epsilon < 0$$

and

$$y_i - \langle w^*, x_i\rangle - b^* + \epsilon > 0$$

then x_i is not a support vector and $\alpha_i = \beta_i = 0$.
Case 2: If

$$y_i - \langle w^*, x_i\rangle - b^* - \epsilon = 0$$

or

$$y_i - \langle w^*, x_i\rangle - b^* + \epsilon = 0$$

then x_i is a support vector.
Notice that we should ignore the case that both $\alpha_i \neq 0$ and $\beta_i \neq 0$, because this means the predicted y_i will locate on both sides of the "margin"(here was represented by $\epsilon > 0$), which is impossible. Thus, the classifier only depends on the support vectors that satisfies the Case2.

3 Problem 3

Part(a): Notice that the SVM classifier is

$$f(x) = \text{sign}\left(\sum_{i=1}^{n} \alpha_i^* y_i k(x, x_i) + b^*\right)$$

where k is the kernal function, where α^* is the solution of

$$\max_{\boldsymbol{\alpha}} -\frac{1}{2}\sum_{i,j=1}^{n} \alpha_i\alpha_j y_i y_j k(x_i, x_j) + \sum_{i=1}^{n}\alpha_i$$

$$\text{st.} \sum_{i=1}^{n}\alpha_i y_i = 0$$

$$0 \leq \alpha_i \leq \frac{C}{n}, \forall i$$

Now suppose we remove a pair of (x_i, y_i), where x_i is not a support vector and then $\alpha_i = 0$, we need to show that with α_{-i}, $L(\alpha_{-i}^*) = L(\alpha^*)$ from maximize the problem of

$$\max_{\boldsymbol{\alpha}_{-i}} -\frac{1}{2}\sum_{k,j=1,k\neq i,j\neq i}^{n} \alpha_j\alpha_k y_j y_k k(x_j, x_k) + \sum_{j=1,j\neq i}^{n}\alpha_i$$

$$\text{st.} \sum_{j=1,j\neq i}^{n}\alpha_j y_j = 0$$

$$0 \leq \alpha_j \leq \frac{C}{n}, \forall j \neq i$$

First notice that $L(\alpha^*_{-i}) > L(\alpha^*)$ is impossible because we know that x_i is not a support vector, if we add back the x_i will result in $L(\alpha^*_i) > L(\alpha^*)$(because we add back many 0 terms), which contradict with the fact that $\boldsymbol{\alpha}$ is the optimal. Similarly, $L(\alpha^*_{-i}) < L(\alpha^*)$ is impossible because this will lead to $L(\alpha^*) < L(\alpha^*)$, which is also a contradiction. Thus, we show that α^*_{-i} must solve the optimization problem and reaches the same objective value as we get from original optimization problem, that is $L(\alpha^*_{-i}) = L(\alpha^*)$, which means, we cannot find another α with dimension of $n-1$ such that $L(\alpha) > L(\alpha^*)$ as we only remove a non-support vector, otherwise, we will contradict with the fact that α^* is optimal.

Part(b): We are given $\alpha_i = 0$, this means

$$f(x) = \text{sign}\left(\sum_{j=1}^{n} a_j^* k(x_j, x) + b^*\right) = \text{sign}\left(\sum_{j=1, j\neq i}^{n} \alpha_j^* y_j k(x_j, x) + b^*_{-i}\right)$$

because

$$\sum_{j=1}^{n} a_j^* k(x_j, x) = \sum_{j=1, j\neq i}^{n} \alpha_j^* y_j k(x_j, x) + \alpha_i^* y_i k(x_i, x)$$

and

$$b^* = y_j - \sum_{k=1}^{n} \alpha_k^* y_k k(x_j, x_k) = y_j - \sum_{k=1, k\neq i}^{n} \alpha_k^* y_k k(x_j, x_k) - \alpha_i y_i k(x_j, x_i)$$

given $\alpha_i = 0$ and for any j such that $0 < \alpha_j^* < \frac{C}{n}$. This completes the proof.