
Machine Learning

Answer Sheet for Homework 5

Da-Min HUANG

R04942045

Graduate Institute of Communication Engineering, National Taiwan University

December 9, 2015

Problem 1

The hard-margin support vector machine is with $d + 1$ variables. For soft-margin support vector machine, there are N more variables ξ_n , $1 \leq n \leq N$.

So soft-margin support vector machine is a quadratic programming problem with $N + d + 1$ variables.

□

Problem 2

I wrote a `Q02.py` to help me get the answer. By using Python package `cvxopt`^[2], with

$$\mathbf{z} = \begin{bmatrix} 1 & -2 \\ 4 & -5 \\ 4 & -1 \\ 5 & -2 \\ 7 & -7 \\ 7 & 1 \\ 7 & 1 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} -1 \\ -1 \\ -1 \\ +1 \\ +1 \\ +1 \\ +1 \end{bmatrix} \quad (1)$$

and

$$\mathbf{Q} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \mathbf{p} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad (2)$$

$$\mathbf{A}^T = \begin{bmatrix} -1 & -1 & 2 \\ -1 & -4 & 5 \\ -1 & -4 & 1 \\ 1 & 5 & -2 \\ 1 & 7 & -7 \\ 1 & 7 & 1 \\ 1 & 7 & 1 \end{bmatrix}, \mathbf{c} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad (3)$$

To use this package, I gave `solvers.qp(Q, p, -AT, -c)` and got

$$b = -9, \mathbf{w} = [2, 0] \quad (4)$$

So the hyperplane is

$$2z_1 - 9 = 0 \Rightarrow z_1 = 4.5 \quad (5)$$

□

Problem 3

I wrote a `Q03.py` to help me get the answer. By using Python package `cvxopt`, with

$$\mathbf{Q} = \begin{bmatrix} 4 & 1 & 1 & 0 & -1 & -1 & -1 \\ 1 & 4 & 0 & -1 & -9 & -1 & -1 \\ 1 & 0 & 4 & -1 & -1 & -9 & -1 \\ 0 & -1 & -1 & 4 & 1 & 1 & 9 \\ -1 & -9 & -1 & 1 & 25 & 9 & 1 \\ -1 & -1 & -9 & 1 & 9 & 25 & 1 \\ -1 & -1 & -1 & 9 & 1 & 1 & 25 \end{bmatrix}, \mathbf{p} = \begin{bmatrix} -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \end{bmatrix}, \quad (6)$$

$$-\mathbf{A}^T = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix}, \mathbf{c} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (7)$$

with

$$\mathbf{G} = \mathbf{y}^T = \begin{bmatrix} -1 & -1 & -1 & 1 & 1 & 1 & 1 \end{bmatrix} \text{ and } h = 0 \quad (8)$$

and To use this package, I gave `solvers.qp(Q, p, -AT, c, G, h)` and got

$$\alpha = [4.32 \times 10^{-9} \approx 0, 0.704, 0.704, 0.889, 0.259, 0.259, 5.27 \times 10^{-10} \approx 0] \quad (9)$$

where `cvxopt` needs conditions

$$-\mathbf{A}^T \boldsymbol{\alpha} \preceq \mathbf{c} \text{ and } \mathbf{G} \boldsymbol{\alpha} = h \quad (10)$$

Support vectors are the corresponding $\alpha_i \neq 0$, so $\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5$ and \mathbf{x}_6 are support vectors. \square

Problem 4

I wrote a `Q04.py` to help me get the answer. By using python package `sympy` and

$$\mathbf{w} = \sum_{n=1}^N \alpha_n y_n K \left(\mathbf{x}_n, \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right) + b \quad (11)$$

$$b = y_s - \sum_{n=1}^N \alpha_n y_n K(\mathbf{x}_n, \mathbf{x}_s) \quad (12)$$

we have

$$\mathbf{w} = \frac{1}{9} (8x_1^2 - 16x_1 + 6x_2^2 - 15) \quad (13)$$

\square

Problem 5

Since kernel function $K(\mathbf{x}, \mathbf{x}') = (1 + \mathbf{x}^T \mathbf{x}')^2$ is different from $\mathbf{z} = (\phi(\mathbf{x}), \phi(\mathbf{x}'))$, the curves should be different in the \mathcal{X} space. \square

Problem 6

Since $\|\mathbf{x}_n - \mathbf{c}\|^2 \leq R^2, \forall n$, the constraint to maximize is

$$\|\mathbf{x}_n - \mathbf{c}\|^2 - R^2 \leq 0 \quad (14)$$

so $L(R, \mathbf{c}, \boldsymbol{\lambda})$ is

$$L(R, \mathbf{c}, \boldsymbol{\lambda}) = R^2 + \sum_{n=1}^N \lambda_n (\|\mathbf{x}_n - \mathbf{c}\|^2 - R^2) \quad (15)$$

□

Problem 7

At the optimal $(R, \mathbf{c}, \boldsymbol{\lambda})$,

$$\frac{\partial L}{\partial R} = 2R - 2R \sum_{n=1}^N \lambda_n = 0 \Rightarrow \sum_{n=1}^N \lambda_n = 1 \text{ or } R = 0 \quad (16)$$

$$\frac{\partial L}{\partial \mathbf{c}} = 2 \sum_{n=1}^N \lambda_n (\mathbf{c} - \mathbf{x}_n) = \mathbf{0} \Rightarrow \mathbf{c} = \left(\sum_{n=1}^N \lambda_n \mathbf{x}_n \right) / \left(\sum_{n=1}^N \lambda_n \right) \text{ if } \sum_{n=1}^N \lambda_n \neq 0 \quad (17)$$

So the KKT conditions are

1. primal feasible: $\|\mathbf{x}_n - \mathbf{c}\|^2 \leq R^2$.
2. dual feasible: $\lambda_n \geq 0$.
3. dual-inner optimal: if $R \neq 0$, $\sum_{n=1}^N \lambda_n = 1$ and $\mathbf{c} = \sum_{n=1}^N \lambda_n \mathbf{x}_n$.
4. primal-inner optimal: $\lambda_n (\|\mathbf{x}_n - \mathbf{c}\|^2 - R^2) = 0$.

□

Problem 8

From Problem 6, we have

$$L(R, \mathbf{c}, \boldsymbol{\lambda}) = R^2 + \sum_{n=1}^N \lambda_n (\|\mathbf{x}_n - \mathbf{c}\|^2 - R^2) = R^2 + \sum_{n=1}^N \lambda_n \|\mathbf{x}_n - \mathbf{c}\|^2 - \sum_{n=1}^N \lambda_n R^2 \quad (18)$$

$$= R^2 - R^2 + \sum_{n=1}^N \lambda_n \|\mathbf{x}_n - \mathbf{c}\|^2 = \sum_{n=1}^N \lambda_n \|\mathbf{x}_n - \mathbf{c}\|^2 \quad (19)$$

where $\sum_{n=1}^N \lambda_n = 1$ since $R \neq 0$.

Also, from (17), we have $\mathbf{c} = \sum_{n=1}^N \lambda_n \mathbf{x}_n$. Hence

$$\text{Objective}(\boldsymbol{\lambda}) = \sum_{n=1}^N \lambda_n \left\| \mathbf{x}_n - \sum_{m=1}^N \lambda_m \mathbf{x}_m \right\|^2 \quad (20)$$

□

Problem 9

We have

$$\sum_{n=1}^N \lambda_n \|\mathbf{x}_n - \mathbf{c}\|^2 = \sum_{n=1}^N \lambda_n (\mathbf{x}_n^T \mathbf{x}_n - \mathbf{x}_n^T \mathbf{c} - \mathbf{c}^T \mathbf{x}_n + \mathbf{c}^T \mathbf{c}) \quad (21)$$

$$= \sum_{n=1}^N \lambda_n \left(\mathbf{x}_n^T \mathbf{x}_n - \mathbf{x}_n^T \sum_{m=1}^N \lambda_m \mathbf{x}_m - \left(\sum_{m=1}^N \lambda_m \mathbf{x}_m \right)^T \mathbf{x}_n + \left\| \sum_{m=1}^N \lambda_m \mathbf{x}_m \right\|^2 \right) \quad (22)$$

So

$$\sum_{n=1}^N \lambda_n \|\phi(\mathbf{x}_n) - \phi(\mathbf{c})\|^2 \quad (23)$$

$$= \sum_{n=1}^N \lambda_n K(\mathbf{x}_n, \mathbf{x}_n) - 2 \sum_{n=1}^N \sum_{m=1}^N \lambda_n \lambda_m K(\mathbf{x}_n, \mathbf{x}_m) + \sum_{n=1}^N \sum_{m=1}^N \lambda_n \lambda_m K(\mathbf{x}_n, \mathbf{x}_m) \quad (24)$$

$$= \sum_{n=1}^N \lambda_n K(\mathbf{x}_n, \mathbf{x}_n) - \sum_{n=1}^N \sum_{m=1}^N \lambda_n \lambda_m K(\mathbf{x}_n, \mathbf{x}_m) \quad (25)$$

□

Problem 10

From primal-inner optimal condition, pick some $\lambda_i > 0$, we have

$$\|\mathbf{x}_i - \mathbf{c}\|^2 = R^2 \quad (26)$$

so

$$R^2 = \mathbf{x}_i^T \mathbf{x}_i - \mathbf{x}_i^T \sum_{m=1}^N \lambda_m \mathbf{x}_m - \left(\sum_{m=1}^N \lambda_m \mathbf{x}_m \right)^T \mathbf{x}_i + \sum_{n=1}^N \sum_{m=1}^N \lambda_n \lambda_m \mathbf{x}_n^T \mathbf{x}_m \quad (27)$$

$$= K(\mathbf{x}_i, \mathbf{x}_i) - 2 \sum_{m=1}^N \lambda_m K(\mathbf{x}_i, \mathbf{x}_m) + \sum_{n=1}^N \sum_{m=1}^N \lambda_n \lambda_m K(\mathbf{x}_n, \mathbf{x}_m) \quad (28)$$

$$\Rightarrow R = \sqrt{K(\mathbf{x}_i, \mathbf{x}_i) - 2 \sum_{m=1}^N \lambda_m K(\mathbf{x}_i, \mathbf{x}_m) + \sum_{n=1}^N \sum_{m=1}^N \lambda_n \lambda_m K(\mathbf{x}_n, \mathbf{x}_m)} \quad (29)$$

where $R > 0$.

□

Problem 11

Claim: Let $\tilde{\mathbf{w}} = \begin{bmatrix} \mathbf{w} \\ \sqrt{2C} \cdot y_1 \xi_1 \\ \sqrt{2C} \cdot y_2 \xi_2 \\ \vdots \\ \sqrt{2C} \cdot y_N \xi_N \end{bmatrix}$ and $\tilde{\mathbf{x}}_n = \begin{bmatrix} \mathbf{x}_n \\ v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix}$, where $v_i = \frac{1}{\sqrt{2C}} \mathbb{I}[i = n]$.

Proof of Claim:

First, we have

$$\frac{1}{2} \tilde{\mathbf{w}}^T \tilde{\mathbf{w}} = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{n=1}^N y_n^2 \xi_n^2 = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{n=1}^N \xi_n^2 \quad (30)$$

where $y_n^2 = 1$ due to $y_n \in \{+1, -1\}$.

And (P_2) can be rewritten as

$$\min_{\tilde{\mathbf{w}}, b, \xi} \left(\frac{1}{2} \tilde{\mathbf{w}}^T \tilde{\mathbf{w}} \right) \quad (31)$$

Then we have

$$\begin{aligned} \tilde{\mathbf{w}}^T \tilde{\mathbf{x}}_n &= \tilde{\mathbf{w}}^T \begin{bmatrix} \mathbf{x}_n \\ v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix} = \tilde{\mathbf{w}}^T \begin{bmatrix} \mathbf{x}_n \\ 0 \\ \vdots \\ 1 \\ \underbrace{\frac{1}{\sqrt{2C}}}_{i=n} \\ \vdots \\ 0 \end{bmatrix} \\ &= \mathbf{w}^T \mathbf{x}_n + 0 + \cdots + y_n \xi_n + \cdots + 0 = \mathbf{w}^T \mathbf{x}_n + y_n \xi_n \end{aligned} \quad (32)$$

$$\quad (33)$$

So

$$y_n (\mathbf{w}^T \mathbf{x}_n + b) \geq 1 - \xi_n = 1 - y_n^2 \xi_n \quad (34)$$

$$\Rightarrow y_n (\mathbf{w}^T \mathbf{x}_n + b) + y_n^2 \xi_n = y_n (\mathbf{w}^T \mathbf{x}_n + y_n \xi_n + b) = y_n (\tilde{\mathbf{w}}^T \tilde{\mathbf{x}}_n + b) \geq 1 \quad (35)$$

Hence, (P_2) is equivalent to a linear hard-margin support vector machine primal problem. \square

Problem 12

Claim: $K(\mathbf{x}, \mathbf{x}') = K_1(\mathbf{x}, \mathbf{x}') + K_2(\mathbf{x}, \mathbf{x}')$ and $K(\mathbf{x}, \mathbf{x}') = K_1(\mathbf{x}, \mathbf{x}') \cdot K_2(\mathbf{x}, \mathbf{x}')$ are always valid kernels.

Proof of Calim:

$$1. K(\mathbf{x}, \mathbf{x}') = K_1(\mathbf{x}, \mathbf{x}') + K_2(\mathbf{x}, \mathbf{x}')$$

Consider Mercer's conditions,

(a) Symmetric

Since K_1 and K_2 are valid kernel, both of them are symmetric. So $K_1 + K_2$ must be symmetry.

(b) K is positive semi-definite

Consider any vector \mathbf{v} , we have

$$\mathbf{v}^T K_1 \mathbf{v} = \left[\sum_{i=1}^N v_i \phi_1(\mathbf{x}_i) \phi_1(\mathbf{x}'_1) \cdots \sum_{i=1}^N v_i \phi_1(\mathbf{x}_i) \phi_1(\mathbf{x}'_N) \right] \mathbf{v} \quad (36)$$

$$= \sum_{j=1}^N \sum_{i=1}^N v_i \phi_1(\mathbf{x}_i) \phi_1(\mathbf{x}'_j) v_j \geq 0 \quad (37)$$

$$\mathbf{v}^T K_2 \mathbf{v} = \left[\sum_{i=1}^N v_i \phi_2(\mathbf{x}_i) \phi_2(\mathbf{x}'_1) \cdots \sum_{i=1}^N v_i \phi_2(\mathbf{x}_i) \phi_2(\mathbf{x}'_N) \right] \mathbf{v} \quad (38)$$

$$= \sum_{j=1}^N \sum_{i=1}^N v_i \phi_2(\mathbf{x}_i) \phi_2(\mathbf{x}'_j) v_j \geq 0 \quad (39)$$

$$\mathbf{v}^T K \mathbf{v} = \mathbf{v}^T (K_1 + K_2) \mathbf{v} = \sum_{j=1}^N \sum_{i=1}^N v_i (\phi_1(\mathbf{x}_i) \phi_1(\mathbf{x}'_j) + \phi_2(\mathbf{x}_i) \phi_2(\mathbf{x}'_j)) v_j \quad (40)$$

$$= \mathbf{v}^T K_1 \mathbf{v} + \mathbf{v}^T K_2 \mathbf{v} \geq 0 \quad (41)$$

Hence, K is positive semi-definite.

By satisfying the Mercer's conditions, K is a valid kernel.

$$2. K(\mathbf{x}, \mathbf{x}') = K_1(\mathbf{x}, \mathbf{x}') \cdot K_2(\mathbf{x}, \mathbf{x}')$$

Similarly, K is symmetry since $K_m(\mathbf{x}_i, \mathbf{x}'_j) = K_m(\mathbf{x}_j, \mathbf{x}'_i)$, for $m = 1$ or 2 . Then

$$K(\mathbf{x}_i, \mathbf{x}'_j) = K_1(\mathbf{x}_i, \mathbf{x}'_j) K_2(\mathbf{x}_i, \mathbf{x}'_j) = K_1(\mathbf{x}_j, \mathbf{x}'_i) K_2(\mathbf{x}_j, \mathbf{x}'_i) = K(\mathbf{x}_j, \mathbf{x}'_i) \quad (42)$$

Applying Cholesky decomposition, rewrite

$$K_1 = A^T A = \sum_{k=1}^N a_{ik} a_{jk} \text{ and } K_2 = B^T B = \sum_{k=1}^N b_{ik} b_{jk} \quad (43)$$

then for any \mathbf{v} , we have

$$\mathbf{v}^T K \mathbf{v} = \sum_{j=1}^N \sum_{i=1}^N v_i \left(\sum_{k=1}^N a_{ik} a_{jk} \right) \left(\sum_{\ell=1}^N b_{i\ell} b_{j\ell} \right) v_j \quad (44)$$

$$= \sum_{k,\ell=1}^N \sum_{i,j=1}^N v_i v_j a_{ik} a_{jk} b_{i\ell} b_{j\ell} \quad (45)$$

$$= \sum_{k,\ell=1}^N \left(\sum_{i=1}^N v_i a_{ik} b_{i\ell} \right) \left(\sum_{j=1}^N v_j a_{jk} b_{j\ell} \right) \quad (46)$$

$$= \sum_{k,\ell=1}^N \left(\sum_{i=1}^N v_i a_{ik} b_{i\ell} \right)^2 \geq 0 \quad (47)$$

Hence, K is positive semi-definite.

By satisfying the Mercer's conditions, K is a valid kernel.

□

Problem 13

Claim: $K(\mathbf{x}, \mathbf{x}') = 1126 \cdot K_1(\mathbf{x}, \mathbf{x}')$ and $K(\mathbf{x}, \mathbf{x}') = (1 - K_1(\mathbf{x}, \mathbf{x}'))^{-1}$ are always valid kernels.

Proof of Claim:

$$1. K(\mathbf{x}, \mathbf{x}') = 1126 \cdot K_1(\mathbf{x}, \mathbf{x}')$$

Consider Mercer's conditions,

(a) Symmetric

Since K_1 is valid kernel, K_1 is symmetric. So K must be symmetric.

(b) K is positive semi-definite

Consider any vector \mathbf{v} , we have

$$\mathbf{v}^T K \mathbf{v} = \sum_{j=1}^N \sum_{i=1}^N v_i (1126 K_1(\mathbf{x}_i, \mathbf{x}'_j)) v_j = 1126 \sum_{j=1}^N \sum_{i=1}^N v_i K_1(\mathbf{x}_i, \mathbf{x}'_j) v_j \quad (48)$$

Since $\sum_{j=1}^N \sum_{i=1}^N v_i K_1(\mathbf{x}_i, \mathbf{x}'_j) v_j \geq 0$, we have

$$1126 \sum_{j=1}^N \sum_{i=1}^N v_i K_1(\mathbf{x}_i, \mathbf{x}'_j) v_j \geq 0 \quad (49)$$

where $\sum_{j=1}^N \sum_{i=1}^N v_i K_1(\mathbf{x}_i, \mathbf{x}'_j) v_j \geq 0$ due to K_1 is positive semi-definite.

Hence, K is positive semi-definite. By satisfying the Mercer's conditions, K is a valid kernel.

2. $K(\mathbf{x}, \mathbf{x}') = (1 - K_1(\mathbf{x}, \mathbf{x}'))^{-1}$

Consider Mercer's conditions,

(a) Symmetric

Since K_1 is valid kernel, K_1 is symmetric. So K must be symmetric.

(b) K is positive semi-definite

Consider any vector \mathbf{v} , we have

$$\mathbf{v}^T K \mathbf{v} = \sum_{j=1}^N \sum_{i=1}^N v_i \left(\frac{1}{1 - K_1(\mathbf{x}_i, \mathbf{x}'_j)} \right) v_j \quad (50)$$

Since $\sum_{j=1}^N \sum_{i=1}^N v_i K_1(\mathbf{x}_i, \mathbf{x}'_j) v_j \geq 0$ and $0 < K_1(\mathbf{x}, \mathbf{x}') < 1$, we have

$$\frac{1}{1 - K_1(\mathbf{x}, \mathbf{x}')} > K_1(\mathbf{x}, \mathbf{x}') \quad (51)$$

where $\sum_{j=1}^N \sum_{i=1}^N v_i K_1(\mathbf{x}_i, \mathbf{x}'_j) v_j \geq 0$ due to K_1 is positive semi-definite. So we have

$$\sum_{j=1}^N \sum_{i=1}^N v_i \left(\frac{1}{1 - K_1(\mathbf{x}_i, \mathbf{x}'_j)} \right) v_j \geq \sum_{j=1}^N \sum_{i=1}^N v_i K_1(\mathbf{x}_i, \mathbf{x}'_j) v_j \geq 0 \quad (52)$$

Hence, K is positive semi-definite. By satisfying the Mercer's conditions, K is a valid kernel.

□

Problem 14

Claim: $\tilde{C} = \frac{C}{p}$, $\tilde{\beta}_n = \frac{\beta_n}{p} = \frac{C}{p} - \frac{\alpha_n}{p} = \tilde{C} - \tilde{\alpha}_n$, $\forall n$ for optimal solution.

Proof of Claim:

$$\tilde{g}_{\text{SVM}}(\mathbf{x}) = \text{sign} \left(\sum_{n=1}^N \tilde{\alpha}_n y_n \tilde{K}(\mathbf{x}_n, \mathbf{x}) + b \right) \quad (53)$$

$$= \text{sign} \left(\sum_{n=1}^N \tilde{\alpha}_n y_n (pK(\mathbf{x}_n, \mathbf{x}) + q) + b \right) \quad (54)$$

$$= \text{sign} \left(p \sum_{n=1}^N \tilde{\alpha}_n y_n K(\mathbf{x}_n, \mathbf{x}) + q \sum_{n=1}^N \tilde{\alpha}_n y_n + b \right) \quad (55)$$

$$= \text{sign} \left(p \sum_{n=1}^N (\tilde{C} - \tilde{\beta}_n) y_n K(\mathbf{x}_n, \mathbf{x}) + q \cdot 0 + b \right) \quad (56)$$

$$= \text{sign} \left(\sum_{n=1}^N \left(p \cdot \frac{C}{p} - p\tilde{\beta}_n \right) y_n K(\mathbf{x}_n, \mathbf{x}) + b \right) \quad (57)$$

$$= \text{sign} \left(\sum_{n=1}^N (C - p\tilde{\beta}_n) y_n K(\mathbf{x}_n, \mathbf{x}) + b \right) \quad (58)$$

where $\sum_{n=1}^N \tilde{\alpha}_n y_n = 0$ due to optimal constraint.

Since $\tilde{\beta}_n = \frac{\beta_n}{p}$, then we have

$$\tilde{g}_{\text{SVM}}(\mathbf{x}) = \text{sign} \left(\sum_{n=1}^N (C - \beta_n) y_n K(\mathbf{x}_n, \mathbf{x}) + b \right) \quad (59)$$

$$= \text{sign} \left(\sum_{n=1}^N \alpha_n y_n K(\mathbf{x}_n, \mathbf{x}) + b \right) \quad (60)$$

$$= g_{\text{SVM}}(\mathbf{x}) \quad (61)$$

□

Problem 15

By using `scikit-learn` package, we have

$$\|\mathbf{w}\|_{\text{list}} = [6.021 \times 10^{-5}, 6.019 \times 10^{-3}, 5.713 \times 10^{-1}, 1.133 \times 10^1, 1.309 \times 10^1] \quad (62)$$

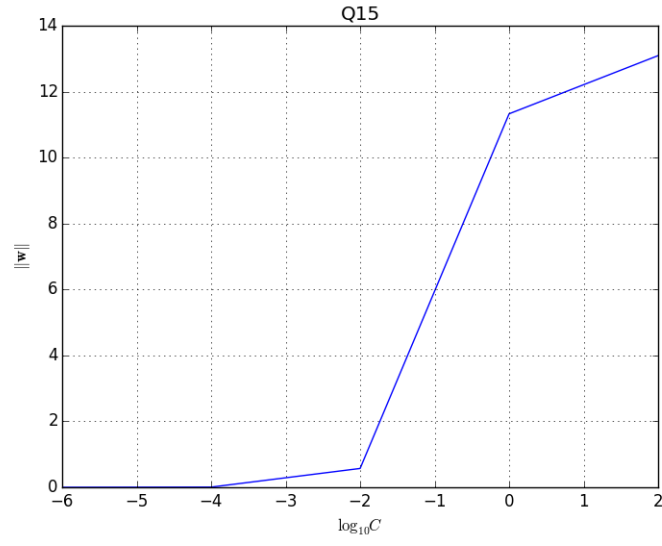


Figure 1: Q15

As C becomes larger, $\|\mathbf{w}\|$ becomes larger, too.

□

Problem 16

All E_{in} versus $\log_{10}(C)$ is 7.434×10^{-2} .

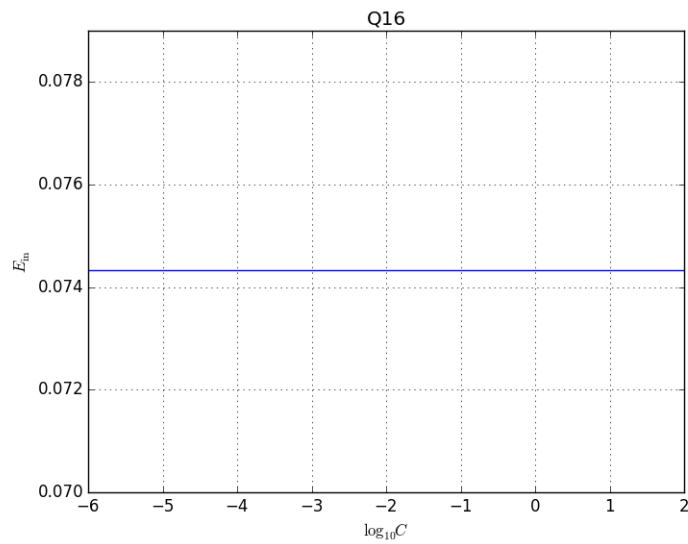


Figure 2: Q16

E_{in} is independent of C .

□

Problem 17

$$\sum_{n=1}^N \alpha_n \text{ list} = [1.084 \times 10^{-3}, 1.084 \times 10^{-1}, 1.084 \times 10^1, 1.084 \times 10^3, 1.084 \times 10^5] \quad (63)$$

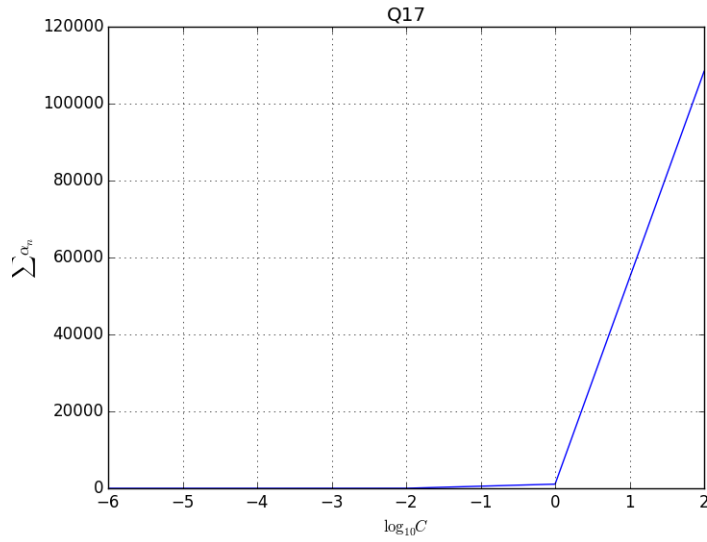


Figure 3: Q17

From this figure, we have

$$\sum_{n=1}^N \alpha_n \propto C \quad (64)$$

□

Problem 18

$$\text{distance list} = [8.237, 8.248 \times 10^{-1}, 1.288 \times 10^{-1}, 8.416 \times 10^{-2}, 4.109 \times 10^{-2}] \quad (65)$$

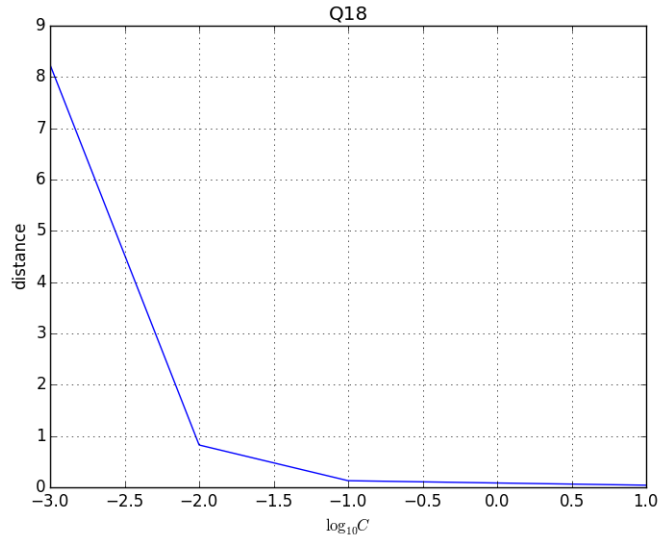


Figure 4: Q18

The distance is strictly decreasing as C becomes larger.

□

Problem 19

$$E_{\text{out list}} = [1.071 \times 10^{-1}, 9.915 \times 10^{-2}, 1.051 \times 10^{-1}, 1.789 \times 10^{-1}, 1.789 \times 10^{-1}] \quad (66)$$

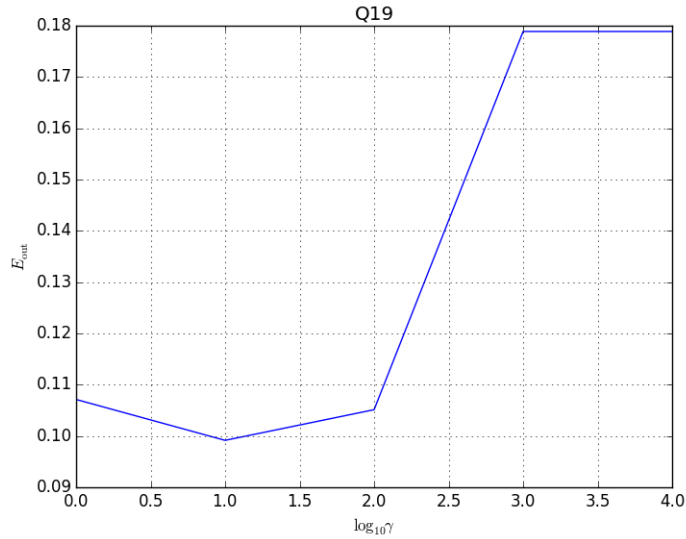


Figure 5: Q19

The minimum occurs at $\gamma = 10$. E_{out} becomes larger as γ becomes larger.

□

Problem 20

The most corresponding γ of $\min E_{\text{val}}$ are 10, 68 times out of 100 repeats and 1: 10 times, 100: 22 times.

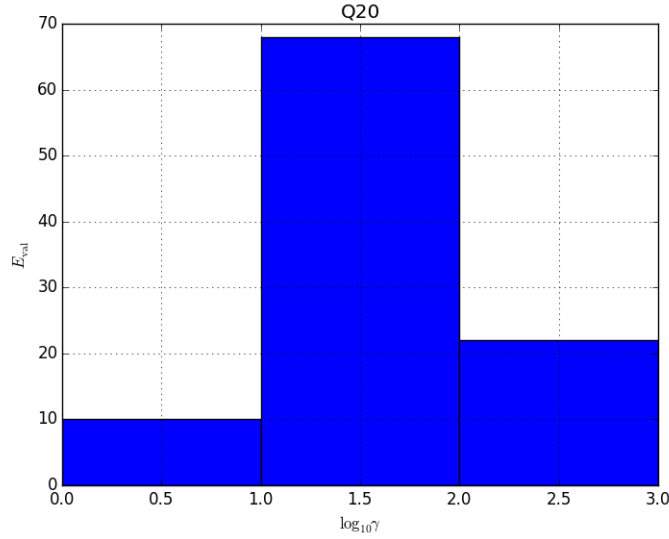


Figure 6: Q20

□

Problem 21

Hard-margin SVM dual:

$$\min_{\alpha} \left(\frac{1}{2} \sum_{n=1}^N \sum_{m=1}^N \alpha_n \alpha_m y_n y_m \mathbf{z}_n^T \mathbf{z}_m - \sum_{n=1}^N \alpha_n \right) \text{ subject to } \sum_{n=1}^N y_n \alpha_n = 0, \alpha_n \geq 0, \forall n \quad (67)$$

then we have

$$\mathcal{L}(b, \mathbf{w}, \alpha, \beta) = \left(\frac{1}{2} \sum_{n=1}^N \sum_{m=1}^N \alpha_n \alpha_m y_n y_m \mathbf{z}_n^T \mathbf{z}_m - \sum_{n=1}^N \alpha_n \right) - \sum_{n=1}^N \beta_n \alpha_n \quad (68)$$

Use $\max_{\beta_n \geq 0, \forall n}$ to eliminate condition $\alpha_n \leq 0$ since if there exists some $\alpha_n < 0$ for some n , then $\max_{\beta_n \geq 0, \forall n} \mathcal{L}(b, \mathbf{w}, \alpha, \beta) \rightarrow \infty$. So the dual SVM is

$$\min_{\alpha} \left(\max_{\beta_n \geq 0, \forall n} \mathcal{L}(b, \mathbf{w}, \alpha, \beta) \right) \quad (69)$$

Hence the Lagrangian dual problem is

$$\min_{\boldsymbol{\alpha}} \left(\max_{\beta_n \geq 0, \forall n} \mathcal{L}(b, \mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \right) \geq \max_{\beta_n \geq 0, \forall n} \left(\min_{\boldsymbol{\alpha}} \mathcal{L}(b, \mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \right) \quad (70)$$

At optimal, we have

$$\frac{\partial \mathcal{L}}{\partial \alpha_i} = 0 = y_i \left(\sum_{n=1}^N \alpha_n y_n \mathbf{z}_n^T \right) \mathbf{z}_i - 1 - \beta_i \Rightarrow \beta_i = y_i \left(\sum_{n=1}^N \alpha_n y_n \mathbf{z}_n^T \right) \mathbf{z}_i - 1 \quad (71)$$

Substitute this into (70), we have

$$\begin{aligned} \max_{\beta_n \geq 0, \forall n} \mathcal{L}(b, \mathbf{w}, \boldsymbol{\alpha}, \boldsymbol{\beta}) &= \max_{\beta_n \geq 0, \forall n} \left(\left(\frac{1}{2} \sum_{n=1}^N \sum_{m=1}^N \alpha_n \alpha_m y_n y_m \mathbf{z}_n^T \mathbf{z}_m - \sum_{n=1}^N \alpha_n \right) \right. \\ &\quad \left. - \sum_{n=1}^N \sum_{m=1}^N \alpha_n \alpha_m y_n y_m \mathbf{z}_n^T \mathbf{z}_m + \sum_{n=1}^N \alpha_n \right) \end{aligned} \quad (72)$$

$$= \max_{\beta_n \geq 0, \forall n} \left(-\frac{1}{2} \sum_{n=1}^N \sum_{m=1}^N \alpha_n \alpha_m y_n y_m \mathbf{z}_n^T \mathbf{z}_m \right) \quad (73)$$

which is

$$\min_{\beta_n \geq 0, \forall n} \left(\frac{1}{2} \sum_{n=1}^N \sum_{m=1}^N \alpha_n \alpha_m y_n y_m \mathbf{z}_n^T \mathbf{z}_m \right) = \min_{\beta_n \geq 0, \forall n} \left(\frac{1}{2} \left\| \sum_{n=1}^N \alpha_n y_n \mathbf{z}_n \right\|^2 \right) \quad (74)$$

$$= \min_{\beta_n \geq 0, \forall n} \left(\frac{1}{2} \|\mathbf{w}\|^2 \right) \quad (75)$$

This is the same as the original problem, but subject to

$$\beta_n \geq 0 \Rightarrow y_n \left(\sum_{m=1}^N \alpha_m y_m \mathbf{z}_m^T \right) \mathbf{z}_n = y_n \mathbf{w}^T \mathbf{z}_n \geq 1 \quad (76)$$

this is different from

$$y_n (\mathbf{w}^T \mathbf{z}_n + b) \geq 1 \quad (77)$$

The reason that b is missing is due to the optimal condition $\sum_{n=1}^N \alpha_n y_n = 0$.

□

Reference

- [1] Lecture Notes by Hsuan-Tien LIN, Department of Computer Science and Information Engineering, National Taiwan University, Taipei 106, Taiwan.

[2] Quadratic Programming with Python and CVXOPT

<https://courses.csail.mit.edu/6.867/wiki/images/a/a7/Qp-cvxopt.pdf>

[3] scikit-learn documentation: 1.4. Support Vector Machines

<http://scikit-learn.org/stable/modules/svm.html#multi-class-classification>