
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

$$\mathbf{G} = \mathbf{y}^T = [-1 \ -1 \ -1 \ 1 \ 1 \ 1 \ 1] \text{ and } h = 0 \quad (8)$$

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

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

\square

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: $K(\mathbf{x}, \mathbf{x}') = K_1(\mathbf{x}, \mathbf{x}') + K_2(\mathbf{x}, \mathbf{x}')$ is always a valid kernel.

Proof of Calim:

Consider Mercer's conditions,

1. Symmetric

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

2. K is positive semi-definite

Consider any vector \mathbf{v} , we have

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

$$= \sum_{j=1}^N \sum_{i=1}^N v_i \phi_1(x_i)^T \phi_1(x'_j) v_j \geq 0 \quad (37)$$

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

$$= \sum_{j=1}^N \sum_{i=1}^N v_i \phi_2(x_i)^T \phi_2(x'_j) v_j \geq 0 \quad (39)$$

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

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

Hence, K is positive semi-definite.

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

□

Problem 13

Claim: $K(\mathbf{x}, \mathbf{x}') = \exp(-K_1(\mathbf{x}, \mathbf{x}'))$ is always a valid kernel.

Proof of Claim:

Consider Mercer's conditions,

1. Symmetric

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

2. K is positive semi-definite

Consider any vector \mathbf{v} , we have

$$\mathbf{v}^T K \mathbf{v} = \mathbf{v}^T \exp(-K_1(\mathbf{x}, \mathbf{x}')) \mathbf{v} = \mathbf{v}^T \sum_{k=1}^{\infty} \frac{1}{k!} (-K_1(\mathbf{x}, \mathbf{x}'))^k \mathbf{v} \quad (42)$$

$$= \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \left(\mathbf{v}^T (K_1(\mathbf{x}, \mathbf{x}'))^k \mathbf{v} \right) \quad (43)$$

where $\mathbf{v}^T (K_1(\mathbf{x}, \mathbf{x}'))^k \mathbf{v} \geq 0$ for all k since K_1 is positive semi-definite. So we have

$$\sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \left(\mathbf{v}^T (K_1(\mathbf{x}, \mathbf{x}'))^k \mathbf{v} \right) \geq 0 \Rightarrow \mathbf{v}^T K \mathbf{v} \geq 0 \quad (44)$$

Hence, K is positive semi-definite.

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

Note: Prove A^k is positive semi-definite if A is positive semi-definite $k \in \mathbb{N}$.

Since A is positive semi-definite, all eigenvalues λ_i of A is non-negative. Hence for any vector \mathbf{v}

$$\mathbf{v}^T A^k \mathbf{v} = \mathbf{v}^T A^{k-1} (A \mathbf{v}) = \mathbf{v}^T A^{k-1} \left(\sum_{i=1}^N \lambda_i v_i \right) = \mathbf{v}^T A^{k-2} \left(\sum_{i=1}^N \lambda_i^2 v_i \right) \quad (45)$$

$$= \dots = \sum_{i=1}^N \lambda_i^k v_i^2 \geq 0 \text{ since } \lambda_i \geq 0, \forall i \quad (46)$$

□

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:

$$\tilde{g}_{\text{SVM}}(\mathbf{x}) = \text{sign} \left(\sum_{n=1}^N \tilde{\alpha}_n y_n \tilde{K}(\mathbf{x}_n, \mathbf{x}) + b \right) \quad (47)$$

$$= \text{sign} \left(\sum_{n=1}^N \tilde{\alpha}_n y_n (pK(\mathbf{x}_n, \mathbf{x}) + q) + b \right) \quad (48)$$

$$= \text{sign} \left(p \sum_{n=1}^N \tilde{\alpha}_n y_n K(\mathbf{x}_n, \mathbf{x}) + q \sum_{n=1}^N \tilde{\alpha}_n y_n + b \right) \quad (49)$$

$$= \text{sign} \left(p \sum_{n=1}^N (\tilde{C} - \tilde{\beta}_n) y_n K(\mathbf{x}_n, \mathbf{x}) + q \cdot 0 + b \right) \quad (50)$$

$$= \text{sign} \left(\sum_{n=1}^N \left(p \cdot \frac{C}{p} - p\tilde{\beta}_n \right) y_n K(\mathbf{x}_n, \mathbf{x}) + b \right) \quad (51)$$

$$= \text{sign} \left(\sum_{n=1}^N (C - p\tilde{\beta}_n) y_n K(\mathbf{x}_n, \mathbf{x}) + b \right) \quad (52)$$

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

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

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

$$= g_{\text{SVM}}(\mathbf{x}) \quad (55)$$

□

Problem 15

By using `scikit-learn` package, we have

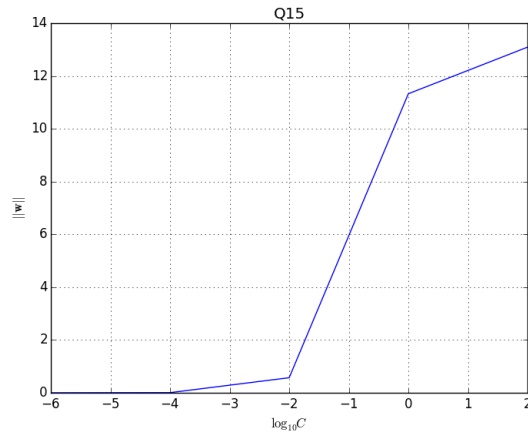


Figure 1: Q15

□

Problem 16

□

Problem 17

□

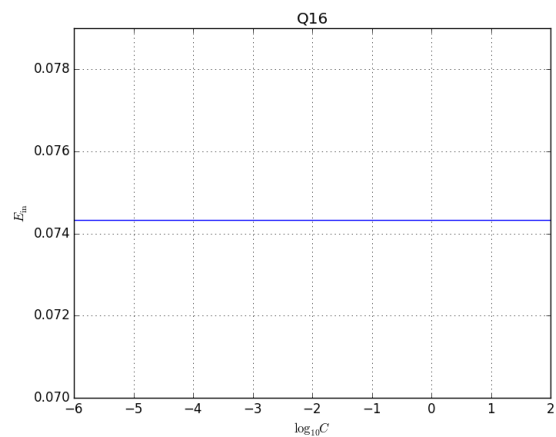


Figure 2: Q16

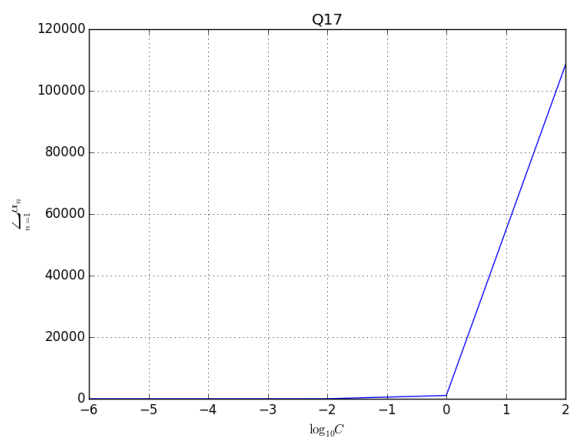


Figure 3: Q17

Problem 18



Problem 19



Problem 20



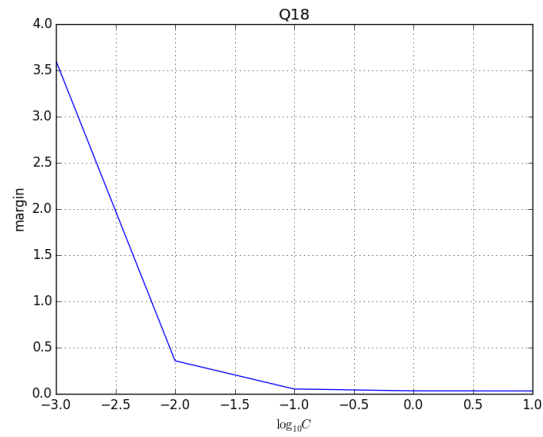


Figure 4: Q18

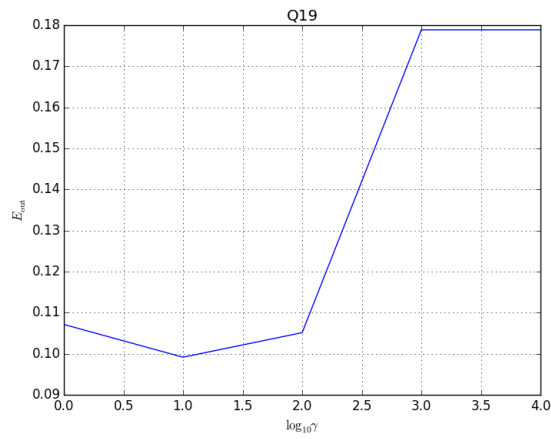


Figure 5: Q19

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>

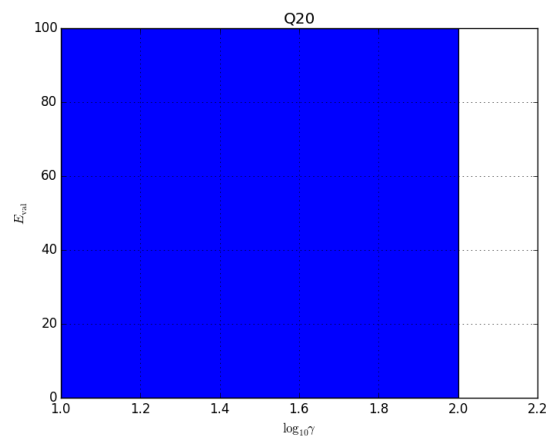


Figure 6: Q20