

Exam 2

CONVEX OPTIMIZATION



April 24, 2023

ITESO UNIVERSIDAD JESUITA DE GUADALAJARA

Alejandro Manuel Aguilar Casillas

Profesor: Juan Diego Sánchez Torres

## REFERENCES

## 1 Exercises …………………………………………………………………………………………………………………………….2

Exercise 1: ……………………………………………………….……………………………………………………………….3

* Clause a……………………………………………………………………………………………………………..
* Clause b……………………………………………………………………………………………………………..
* Clause c……………………………………………………………………………………………………………..
* Clause d……………………………………………………………………………………………………………..

Exercise 2: ……………………..…………………………………………………………………………………………….

* Clause a……………………………………………………………………………………………………………..
* Clause b……………………………………………………………………………………………………………..
* Clause c……………………………………………………………………………………………………………..

Exercise 3……………………………………………………………………………………………………………………..

Exercise 4……………………………………………………………………………………………………………………..

* Clause a………………………………………………………………………………………………………….…
* Clause b………………………………………………………………………………………………………….…
* Clause c………………………………………………………………………………………………………….…
* Clause d………………………………………………………………………………………………………….…

Exercise 5…………………………………………………………………………………………………………………..…

* Clause a………………………………………………………………………………………………………….…
* Clause b……………………………………………………………………………………………………………
* Clause c……………………………………………………………………………………………………………
* Clause d……………………………………………………………………………………………………………
* Clause e………………………………………………………………………………………………………….…
* Clause f……………………………………………………………………………………………………………
* Clause g……………………………………………………………………………………………………………
* Clause h……………………………………………………………………………………………………………

Exercise 6………………………………………………………………………………………………………………………

Exercise 7………………………………………………………………………………………………………………………

Exercise 8………………………………………………………………………………………………………………………

1. (20 pt) Warm-up: Let the following optimization exercises:

(a) (5 pt) Consider the linear optimization problem (linear program)

Prove that the Lagrangian of (1) is For this case, ; then, the dimensions of c and A follow from and . Show that the Lagrange dual function (The Wolfe’s dual) of (1) is

and then, write the complete dual optimization problem. This dual problem is also linear program, but with m variables.

(b) (5 pt) Consider the quadratic optimization problem

with Q ≻ 0 and symmetric. Note that the Lagrangian of (2) is For this case, ; then, the dimensions of follow from and . Show that the Lagrange dual function (The Wolfe’s dual) of (2) is

and then, write the complete dual optimization problem.

(c) (5 pt) Consider the quadratic optimization problem

with Q ≻ 0 and symmetric. Show that, by the KKT conditions, for this convex problem with no inequality constraints, x is a solution if satisfies

for some v. That is, the system (4) provides the optimality conditions for (3).

(d) (5 pt) Consider the quadratic optimization problem

with Q ≻ 0. Find the optimality conditions for (5).

2. (10 pt) Consider the following optimization problem:

where represents a training set with input data , the output data given are slack variables, and the feature maps have the form . Then, the model’s parameters are . Finally, .

Note that the problem (6) can be written as:

Thus, the problem (6) is related to the so-called least squares support vector machines LS-SVM.

(a) Show that the Lagrangian of the problem (6) is given by:

(b) Since the problem has not inequality constraints, the KKT optimality follows directly from the first order conditions provided by the gradient of the Lagrangian . Then, show that:

(c) Define adequate vector variables, such that the optimization problem reduces to a set of linear equations which must be solved for α and b.

3. (10 pt) Consider the following optimization problem:

where is the response (target) variable.

Then conduct an analysis of the problem (7) by applying the steps (a) to (d) of the problem (6) with, possibly, their respective modifications. In the step (d), explain how the new problem (7) is related to the classification problem. Finally, compare KKT matrix system obtained for this case with that of the problem (6).

4. (20 pt) Let be the training data where are deterministic points (fixed design) and an unknown real-valued smooth function and uncorrelated random errors with . The model for regression is given as denotes a potentially infinite dimensional feature map.

Then, consider the cost functions:

• Tikhonov:

• Morozov’s discrepancy principle, where the minimal 2-norm of w realizing a fixed noise level is to be found:

• Ivanov regularization amounts at solving for the best fit with a 2-norm on w smaller than . The following modification is considered in this case

The use of the equality (instead of the inequality) can be motivated in a kernel machine context as these problems are often ill-conditioned and result in solutions on the boundary of the trust region .

Show that:

(a) The conditions for optimality are

|  |  |  |  |
| --- | --- | --- | --- |
| Condition | Tikhonov | Morozov | Ivanov |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

. The kernel-trick is applied as follows: for an appropriate kernel in order to avoid explicit computations in the high dimensional feature space. Let be such that

(b) The Tikhonov conditions result in the following set of linear equations as classical:

(c) Re-organizing the sets of constraints of the Morozov and Ivanov scheme results in the following sets of linear equations where an extra nonlinear constraint relates the Lagrange multiplier ξ with the hyper-parameter

Morozov:

(d) and,

Ivanov :

5. (10 pt) Consider the following optimization problem:

where is the response (target) variable, are slack variables, and c > 0. The remaining variables are defined as in (6).

(a) Explain (with math) how the optimization problem (8) is related to the binary classification problem. That is, deduce the optimization problem related to the linearly separable binary classification problem. After that, reformulate the optimization problem considering the possibility of separating the two sets with nonlinear functions. In addition, introduce slack variables (soft margin) to consider the possibility that some samples are misclassified.

(b) Show that the Lagrangian of the problem (8) is given by:

(c) Calculate the dual cost function (The Wolfe Lagrangian) .

(d) Show the dual problem in the Lagrange multipliers (The Wolfe dual problem) as a quadratic programming formulation.

(e) For the problem (8), derive the KKT conditions:

* + Stationary condition.
  + Primal feasibility condition.
  + Dual feasibility condition.
  + Complementary slackness condition.

Then, compare the result obtained with the requirements of the KKT formulation with that given in the dual problem.

(f) Use the complementary slackness condition to prove the existence of some . Then, confirm or refute the following conditions related to :

If those conditions hold true, rewrite the primal problem (8) and its Lagrangian function in terms of the so-called Hinge loss.

The property of having some is called sparseness.

(g) Present a rigorous method to find the value of b.

(h) Finally, provide the following refinements and explanations:

* Define and write the dual problem in terms of
* Show that the solution vector w has an expansion in terms of the training vectors with Then, explain the effect of the sparseness on w.
* Those Lagrange multipliers such that are called support values, and those vectors corresponding to the support values in the expansion for the solution vector w are called support vectors. Firstly, notice that, by definition, only the support values and vectors are relevant to provide a solution for the optimization problem (8).

6. (10 pt) Note that the problem (8) defines a support vector machine for solving binary classification problems. In this case, state appropriately (as a convex optimization problem) the analogous problem for the regression case. After that, present the solution to the proposed optimization problem using a solution which is analogous to the one used in the classification problem (8).

7. (10 pt) Consider the following optimization problem:

where . The remaining variables are defined as usual.

For the problem (9), derive the KKT conditions:

* Stationary condition.
* Primal feasibility condition.
* Dual feasibility condition.
* Complementary slackness condition.

Use the previous conditions to present a complete and detailed development of the regression model related to the optimization problem (9).

8. (10 pt) Note that the problem (9) defines a support vector machine for solving regression problems. In this case, state appropriately (as a convex optimization problem) the analogous problem for the binary classification case. After that, present the solution to the proposed optimization problem using a solution which is analogous to the one used in the regression problem (9).