FM3817B Assignment 2

Created by Yifan Li

Released: February 11, 2025

All questions are finalized.

Due: March 4, 2025, 11:55 pm

General information Full scores in total: 60 points (pts). Question(s) with bonus points (10 pts): If you do it correctly, you will get the bonus pts (until reach the full scores). Otherwise, you will not lose any points. Hence, feel free to try it out and play with them! Enjoy.

More questions will be added progressively as the course advances. This assignment will be finalized by February 25, 2025. Students are strongly encouraged to begin working on it early, rather than waiting until the last day.

Submission Rules All submissions must be made via Gradescope. You are required to carefully assign each page to its corresponding question to ensure proper grading. Failure to follow these guidelines will result in a deduction of points. Typical penalties apply in the following cases:

a. Submission is not in PDF format.

b. Questions with no assigned pages.

Submissions must be uploaded as a **single PDF file**. You may create this file electronically using a tablet or software such as Microsoft Word, LaTeXetc. If necessary, you can generate a PDF using the following methods:

a. If you write your derivations by hand, scan the pages into a single PDF file.

b. If you have images of your work, insert them into a Word document and save it as a PDF.

If you experience difficulties formatting your submission, seek assistance from the TA during office hours as soon as possible.

Important rule It is your responsibility to allocate some time to format your submission prior to the due time and ensure that you can work with Gradescope properly. Late submissions are NOT accepted.

Other policies You are allowed to discuss with your classmates. However, each student must submit their *own work*. Scholastic offences are taken seriously, and students are directed to read the appropriate policy, specifically, the definition of what constitutes a Scholastic Offence, at this link. Please turn to the course outline for more details on policies related to assignments.

Important Tip The primary goal of the assignment is to help you reinforce and apply the concepts discussed in the lecture. The most effective approach is to *first review the course material*, and then attempt the questions *independently*, with minimal external assistance. This process will enable you to practice and internalize the concepts more thoroughly. (At least, this is precisely what will be expected during the exam.)

Question 1 (15 pts). We are going to explore an important application of optimization in statistical learning, the so-called *ridge regression*.

Background and Motivation

Suppose we have the dataset (x_i, y_i) with $i = 1, 2, \ldots, n$. In ordinary least squares (OLS) regression, we estimate β_0 and β_1 by minimizing the sum of squared errors:

$$\min_{\beta_0,\beta_1} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2. \tag{1}$$

However, when predictor variables are highly correlated or when we have limited data, OLS estimates can become *unstable* and have *high variance*, meaning small changes in the data can lead to large changes in β_0 and β_1. This can cause overfitting, where the model performs well on training data but poorly on new data.

To address this, *ridge regression* adds a constraint on β_0 and β_1 to prevent them from becoming too large. Instead of solving the standard OLS problem, we solve:

$$\min_{\beta_0,\beta_1} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2 \quad \text{subject to} \quad \beta_0^2 + \beta_1^2 \leq t. \tag{2}$$

This formulation ensures that the estimated coefficients do not grow excessively, helping to reduce overfitting.

Assignment Tasks:

a. The OLS regression problem (1) can be treated as unconstrained optimization. Find the optimal β_0 and β_1.[1]

b. Write down the Lagrangian function by introducing the multiplier λ for this constrained optimization problem (2).[2]

c. Suppose the NDCQ condition holds for (2). Find the optimal β_0 and β_1 (in terms of the data and λ) by using the Lagrangian method. [3]

d. Compare your optimal β_0 and β_1 in OLS regression versus ridge regression. Explain why ridge regression *shrinks* the estimates of β_0 and β_1.

Question 2 (10 pts). Portfolio optimization is an important subject in financial modeling and has a long history. One of the foundational stones of the portfolio theory is the Markowitz model. We are going to cover this in detail in later part of the course but we can still do some preview using what we have learned before. Suppose there are three stocks and the allocated proportion is (x, y, z). Since they are proportions, the sum of them is equal to 1. The required mean return of the portfolio is *equal to* 4. From the estimation results of the historical data, the mean return vector of each share of stock is $(1, 10, 6)$ and their variance are $(1, 4, 2)$ (we can see that the stock with higher mean is also more volatile). The correlation between the first and second stock is 0.125. Then the variance of the portfolio with allocation (x, y, z) can be written as

$$x^2 + 0.5xy + 4y^2 + 2z^2.$$

a. We want to minimize the variance of the portfolio to obtain our *optimal portfolio*. In the course, we have mainly focused on maximization problem. How should we modify the objective function such that it becomes an maximization problem?
b. Please write down the equality constraints.
c. Obtain the optimal allocation by applying what we have learned in constrained optimization problem. [4]
d. Translate the problem into an univariate unconstrained optimization and solve it.
e. Suppose the required mean return of the portfolio is increased to 4.5. Approximate the change in the *minimum variance* of the portfolio.
f. Suppose the mean return of the third stock is increased to 6.5. Approximate the change in the *minimum variance* of the portfolio.

Question 3 (6 pts)**.** As a continuation of the previous problem, suppose we require the mean return of the portfolio to be *at least* 4. We also require there is no short-selling (we can only purchase stocks).

a. Write down the *inequality* constraints. [5]
b. Write the Kuhn-Tucker formulation of the problem (Lagrangian and the first-order conditions). [6]
c. (Bonus 4 pts) Suppose NDCQ holds for this problem. Solve this problem using Kuhn-Tucker formulation. [7]

Question 4 (4 pts)**.** Sketch the following subsets of $\mathbb{R}^2$. Are they convex or not? (Yes or No) Do they have extreme points? (If yes, specify them and show your computation steps).

a. The system:

$$\begin{cases} x + 2y & \leq 1 \\ y - 2x & \geq -2 \\ x + 5 & \geqslant 0 \end{cases}$$

b. The system:

$$\begin{cases} x + y & \geq 0 \\ x - y & \leq 0 \\ x & \geqslant 0 \end{cases}$$

c. (Bonus 3 pts) The system [8]:

$$\begin{cases} x + y & \geq 1 \\ x^2 + y^2 & \leq 1 \end{cases}$$

Question 5 (25 pts)**.** (Conceptual questions)[9] Please briefly explain the meaning of the following concepts (using formula or graphs to augment your explanation if needed) and answer the related question:

a. A hyperplane in $\mathbb{R}^n$. Please sketch $x + y = 1$ in $\mathbb{R}^2$. Further sketch it in $\mathbb{R}^3$.

b. The half-spaces determined by a hyperplane. In $\mathbb{R}^2$, please sketch the two half spaces determined by $2x + y = 4$. Does the line $2x + y = 4$ separate two points $(1, 0)$ and $(0, 5)$? (Yes or No) Why?

c. A hyperplane H separating two sets A and B. In $\mathbb{R}^2$, does the line $2x + y = 4$ separate the two sets (respectively) specified by $x^2 + y^2 \leq 1/4$ and $(x - 2)^2 + (y - 2)^2 \leq 1/4$? (Yes or No)[10] (You will get Bonus 3 pts if analytical reasons are provided.[11])

d. Theorem of the separating hyperplane. Can we remove the condition of "convex sets"? (Yes or No) Why? (If your answer is No, draw a counter example.)

e. A supporting hyperplane H of a set S. What is a tangential hyperplane? Draw two examples: one for illustrating tangential hyperplane, and the other for a supporting hyperplane but is not a tangential one.

f. An extreme point of S. What are the extreme points of the set decided by $x \in [0, \pi]$ and $y \leq \sin(x)$? (Please draw them out. No need to state reasons.)[12]

Notes

[1]Hint: check both the first and second order condition.

[2]Here are some interesting facts for your curiosity: the Lagrangian multiplier λ plays an important role in ridge regression. As we will see later in this question, it is actually a *tuning parameter* that controls how much *shrinkage penalty* we put on the estimation of regression coefficients β_0 and β_1. In practice, we usually involve higher dimension of problems (that is, many more coefficients β_i with $i = 0, 1, 2, \ldots, p$ or more predictors for y). For simplicity of this problem, we only consider $p = 2$. However, if p is large, we usually prefer to choose large λ to put more penalty on models with too many variables to achieve simpler models (which is called better *sparsity* in model).

[3]You do not need to solve for λ in the context of ridge regression, which is specified by the user in practice (through some data techniques like cross-validation). If you are interested, turn to this link for more details.

[4]The main goal here is that you need to write down the system of conditions correctly. There is no need to worry about the non-negativity for now. For the computation, you can use whichever way that is more convenient (manual way, software, calculators, etc.) Please specify the tool you use (e.g. calculator, matrix solver, etc.) to get the results (if you do not solve it manually). It is the same for the computation in the following sub-question (d).

[5]Hint: This question only asks for the *inequality* constraints. There is no need to worry about the constraint that the sum equal to 1 since it is an equality constraint (of course, you will not lose any marks if you include it), but you need to consider this in the next subquestion.

[6]Hint: Be careful with the direction of the inequalities. Note that this problem requires careful thinking because there is an *equality* constraint involved (the sum is equal to 1). Please recall the idea of the Kuhn-Tucker formulation we have mentioned in class (or Section 18.6 in the book by Simon and Blume [1]). Here is a candidate way:

a. First write down the original Lagrangian formulation with the first-order conditions. Do we need non-negativity for the multiplier of an equality constraint? What does the first-order condition look like for an equality constraint?

[1]Simon, C.P. and Blume, L. Mathematics for Economists

b. Recall the relationship between the derivatives of the original and Kuhn-Tucker Lagrangian.

c. Then write down the Kuhn-Tucker Lagrangian with the first-order conditions. Specially speaking, can you write the first-order condition (for the equality constraint) using the derivative of the Kuhn-Tucker Lagrangian?

These guiding questions are designed to help you during the thinking and there is no need to answer all of them if you already know how to do it.

[7]Learning tip: please also pay attention to the difference in solving it compared with the original formulation. This bonus question requires you to do some exploration how to solve a problem with mixed constraints. Although we have not covered it in the lecture, you can directly combine what we have learned in equality and inequality constraints (pay attention to which multipliers need to be ≥ 0) to solve the system. For more details, please turn to the book by Simon and Blume (Section 18.4).

[8]Hint: Recall *carefully* the definition and meanings of extreme points. It is recommended to think about it a more geometric and conceptual way, rather than run through some complicated derivation.

[9]This question is designed to let students have more practice involving examples on the geometric concepts we have prepared for LP. Meanwhile, it is aimed for further enhancing your practice on conceptual questions.

[10]Hint: You may draw a graph to help you find the answer. The graph is optional but recommended to draw as a practice.

[11]Hints: To receive the bonus points, the geometric graphs are not enough. You need to give a more rigorous check. Recall that the distance between two points (x_1, y_1) and (x_2, y_2) in $\mathbb{R}^2$ is $\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$. You can compute the *minimum* distance between a point on $2x + y = 4$ and the center of the circle, to check whether the line cross the circle or not. As another way (optional), you can also use the formula between a point (x_0, y_0) and a line $ax + by + c = 0$:

$$d = \frac{|ax_0 + by_0 + c|}{\sqrt{a^2 + b^2}}.$$

[12]Hints: recall the definition of $\sin(x)$. This course assumes your familiarity with those common functions such as trigonometric functions in calculus.