

Please answer the following questions in complete sentences in a clearly prepared manuscript and submit the solution by the due date on Gradescope.

Remember that this is a graduate class. There may be elements of the problem statements that require you to fill in appropriate assumptions. You are also responsible for determining what evidence to include. An answer alone is rarely sufficient, but neither is an overly verbose description required. Use your judgement to focus your discussion on the most interesting pieces. The answer to “should I include ‘something’ in my solution?” will almost always be: Yes, if you think it helps support your answer.

Problem 0: Homework checklist

- Please identify anyone, whether or not they are in the class, with whom you discussed your homework. This problem is worth 1 point, but on a multiplicative scale.
- Make sure you have included your source-code and prepared your solution according to the most recent Edstem note on homework submissions.

Problem 1: Convexity and least squares

1. Show that $f(x) = \|\mathbf{b} - \mathbf{A}\mathbf{x}\|^2$ is a convex function. Feel free to use the result proved on the last homework.
2. Show that the null-space of a matrix is a convex set. (A convex set satisfies the condition that, for every pair of points in the set, any point on the line joining those points is also in the set.)
3. Chat with an LLM about whether or not a discontinuous function can be convex. It’s going to throw lots of jargon at you. Make a list of all the terms it throws at you and ask it for an explanation of each.

Problem 2: Ridge Regression

The Ridge Regression problem is a variant of least squares:

$$\text{minimize } \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_2^2.$$

This is also known as Tikhonov regularization.

1. Show that this problem always has a unique solution, for *any* \mathbf{A} if $\lambda > 0$, using the theory discussed in class so far.

This property is one aspect of the reason that Ridge Regression is used. It is also a common regularization method that can help avoid overfitting in a regression problem.

2. Use the SVD of \mathbf{A} to characterize the solution as a function of λ .
3. What is the solution when $\lambda \rightarrow \infty$?
What is the solution when $\lambda \rightarrow 0$?
4. Give the solutions of the least squares problem using the sports teams data when $\lambda = 0$ and $\lambda = \infty$. (Here, you will not need the constraint that the

sum of entries is 1.) Any notable differences from the ranking giving in class?

5. Suppose that you only want to regularize two components of the solution, say, x_1 , so that your optimization problem is

$$\text{minimize } \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda x_1^2 + \lambda x_2^2.$$

Show how to adapt your techniques in this problem to accomplish this goal.

Problem 3: Thinking about constraints

Find the minimizer of $f(x, y) = 2x^2 + 3y^2$ where also $y = -x + 4$ (that is, we have added one linear constraint) and justify, as concisely and correctly as you can, that you have found a solution. Do the terms local and global minimizer apply? If so, use one of them to describe your solution. Furthermore, evaluate the gradient of f at your solution. What is notable about the gradient at your solution in comparison with the solution where x, y have no constraints?

Problem 4: Alternate formulations of Least Squares

Consider the constrained least squares problem:

$$\begin{aligned} & \underset{\mathbf{r}, \mathbf{y}}{\text{minimize}} && \|\mathbf{r}\|_2 \\ & \text{subject to} && \mathbf{r} = \mathbf{b} - \mathbf{C}\mathbf{y} \end{aligned}$$

where $\mathbf{C} \in \mathbb{R}^{m \times n}$, $n \leq m$ and rank n .

1. Convert this problem into the standard constrained least squares form.
2. Form the augmented system from the Lagrangian as we did in class.
3. Manipulate this problem to arrive at the normal equations for a least-squares problem: $\mathbf{C}^T \mathbf{C}\mathbf{y} = \mathbf{C}^T \mathbf{b}$.
Discuss any advantages of the systems at intermediate steps.

(Optional) Problem 5: Develop an algorithm

Work with an LLM to develop an algorithm to solve the p -norm least squares problem

$$\text{minimize } \sum_{i=0}^{\gamma} |\gamma - b_i|^p$$

for $p \geq 1$. What happens when $0 < p < 1$? Try and get it to develop a report with plots and an explanation.