

CARNEGIE MELLON UNIVERSITY 07-380

HOMEWORK 3

<https://www.cs.cmu.edu/~07380>

INSTRUCTIONS

- **Format:** Use the provided LaTeX template to write your answers in the appropriate locations within the *.tex files and then compile a pdf for submission. We try to mark these areas with STUDENT SOLUTION HERE comments. Make sure that you don't change the size or location of any of the answer boxes and that your answers are within the dedicated regions for each question/part. If you do not follow this format, we may deduct points.

You may also digitally annotate the pdf. Illegible handwriting will lead to lost points. However, we suggest that you try to do at least some of your work directly in LaTeX.

- **How to submit written component:** Submit to Gradescope a pdf with your answers. Again, make sure your answer boxes are aligned with the original pdf template.
- **Policy:** See the course website for homework policies, including late policy, and academic integrity policies.

Name	
Andrew ID	
Hours to complete all components (nearest hour)	

1 [12 pts] Integer Programming: Bayes the Bat's Portfolio

Stock portfolio investment is a common application of linear and integer programming. Bayes the Bat is interested in investing in two companies, EchoCorp and Sonar Inc. Let x_1 represent the number of shares of EchoCorp stock he purchases and x_2 be the number of shares of Sonar Inc. A good way to hedge his bets is to diversify his portfolio, so he wants at least 2 shares of EchoCorp's stock and at least 2 shares of Sonar Inc's. One share of EchoCorp costs \$1, and one share of Sonar Inc. costs \$2. He wants to spend at most \$10. He also wants to keep his risk below some risk threshold R . Each share of EchoCorp has risk 3, and each share of Sonar Inc. has risk 1. How much of each stock should he buy to maximize profit if EchoCorp has profit \$4 per share, and Sonar Inc. has profit \$1 per share?

1. [2 pts] Write the A matrix, $\mathbf{b}$ vector, and $\mathbf{c}$ vector for this problem in **inequality form** (minimization), as defined in lecture:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^T \mathbf{x} \\ \text{s.t.} \quad & A\mathbf{x} \preceq \mathbf{b} \end{aligned}$$

Assume $\mathbf{x} = [x_1, x_2]^T$ in that order. Use the parameter R where needed.

A

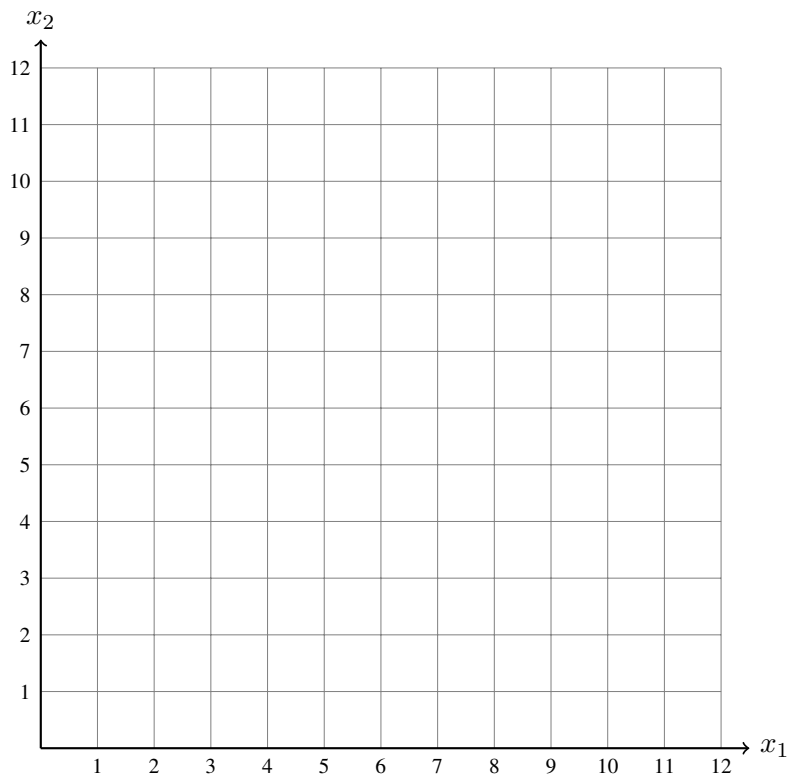
$\mathbf{b}$

$\mathbf{c}$

2. [2 pts] If $R = 13$, what are the corners (x_1, x_2) of the feasible region (the constrained polygon)?

Corners

(Optional) For partial credit on parts (b) and (c), you may show your work on the grid below.



3. [1 pts] Still with $R = 13$, what are the coordinates (x_1, x_2) of the optimal solution to the LP (ignoring any integer requirement)? What is the optimal profit?

LP Solution Coordinates

Profit

4. [4 pts] You should find that the optimal LP corner is not an integer point. On his stock trading site, Bayes must buy whole shares of stock. Run branch and bound (with $R = 13$) to find the optimal integer number of shares of each stock he should buy. Follow these instructions very carefully:

- Each iteration corresponds to one depth of the branch and bound tree. Fill in the appropriate information for each iteration until the branch and bound algorithm finishes and returns. *Use only as many rows as the algorithm needs before it returns.*
- As in lecture, branch on the non-integer coordinate x_i with value v by adding the constraint $x_i \leq \lfloor v \rfloor$ for the left branch and $x_i \geq \lceil v \rceil$ for the right branch. If more than one coordinate is non-integer, branch on x_1 first.
- Specify **ALL** the constraints you add to the problem *in addition to the original set*. For example, if you first branch on $x_2 \leq 10$ and $x_2 \geq 11$, then you'd write for the left branch " $x_2 \leq 10$ " and for the right branch " $x_2 \geq 11$ ". If you then pull $x_2 \geq 11$ off the queue and branch on $x_1 \leq 20$ and $x_1 \geq 21$, the second iteration would have left branch " $x_2 \geq 11, x_1 \leq 20$ " and right branch " $x_2 \geq 11, x_1 \geq 21$ ".
- Give each branch's solution as the LP solution (x_1, x_2) of that branch. If a particular branch is infeasible, write "*infeasible*" in that box.

Iter.	Left branch constraints	Left solution (x_1, x_2)	Right branch constraints	Right solution (x_1, x_2)
1				
2				
3				
4				

What is the optimal integer solution and its profit?

IP Solution Coordinates

Profit

5. **[3 pts]** Bayes decides that he is willing to be riskier in his stock purchases in order to make more money. Do not change any constraints except the value of the risk threshold R . Find the minimum value of R that maximizes the total (integer) profit given the rest of the constraints.

Minimum R

New IP Solution

New Profit

2 [6 pts] Ethical Considerations for Amazon Delivery Routes

The following question is about the article *Amazon's Cost Saving Routing Algorithm Makes Drivers Walk Into Traffic*:

<https://www.vice.com/en/article/amazons-cost-saving-routing-algorithm-makes-drivers-walk-into-traffic/>

1. [2 pts] Give two possible parameters that the Amazon delivery routing algorithm currently optimizes for. Why do you think these are being used?

Your Answer

2. [2 pts] Why are these parameters in conflict with the delivery drivers' needs? Additionally, give two possible parameters that the routing algorithm could add to make it more fair for the drivers.

Your Answer

3. [2 pts] Mike, the anonymous interviewee, gave a lot of information about his route and troubles in the interview. Do you think it would be ethical for Amazon, using their data on delivery routes, to design an algorithm to search for who Mike really was? Why or why not?

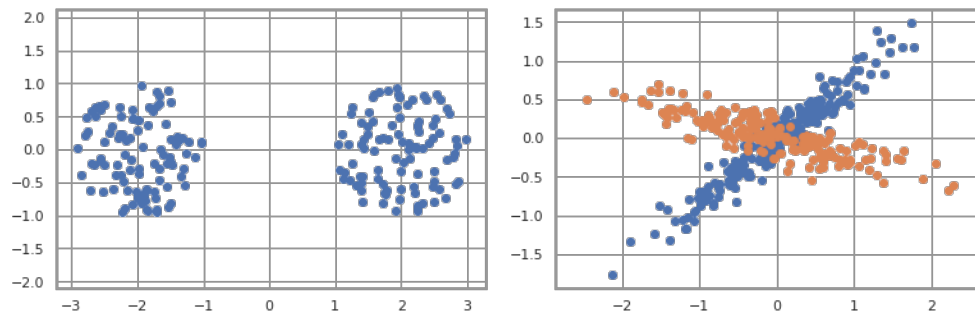
Your Answer

3 [9 pts] PCA

3.1 [2 pts] PCA Warm-Up

1. [2 pts] Principal component analysis is a dimensionality reduction method that projects a dataset into its most variable components. You are given the following 2D datasets, draw the first and second principal components on each plot. Be sure to label which vector is the first principal component and which is the second.

Update plots



3.2 [7 pts] PCA Computation

Given 6 data points in $\mathbb{R}^5$, represented as rows in a 6 x 5 matrix X below:

$$X = \begin{bmatrix} -2 & -2 & -2 & 0 & 0 \\ -2 & -2 & -2 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 0 & 0 & 0 & -2 & -2 \\ 0 & 0 & 0 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

1. [1 pts] Is X centered?

☐ Yes ☐ No

2. [2 pts] Write X in its SVD form (a product of three matrices) where the dimensions of the three matrices are as small as possible such that X may be recovered.

SVD decompositions are not unique. To help with grading (and with the rest of PCA below), write the left matrix with columns that have L2 norm equal to one and the right matrix with rows that have L2 norm equal to one.

You may calculate the values however you like.

Your Answer

3. [1 pts] What is the first principal component of the original data set?

Your Answer

4. [1 pts] If we project the original data set onto 1-D space using the first principal component, what's the variance of the projected data?

Your Answer

Optional space to show your work:

Your Answer

5. [2 pts] For the projected data in the previous part, what is the reconstruction error? Report your answer up to three decimal places.

Your Answer

Optional space to show your work:

Your Answer

4 [12 pts] MAP: Priors and Regularization

In this problem, we are going to have you work through the proof that adding a Laplace prior to probabilistic linear regression is the exact same as adding an L1 regularization to the MSE optimization of linear regression!

Remember our probabilistic linear regression set-up, in which we assume the relationship between response $y \in \mathbb{R}$ and an input $\mathbf{x} \in \mathbb{R}^M$ is linear with zero-mean Gaussian noise added. Note that in this problem for simplicity, we are going to assume that there is no bias term at all.

$$y = \mathbf{w}^T \mathbf{x} + \epsilon \text{ with } \epsilon \sim \mathcal{N}(0, \sigma^2)$$

or more concisely, $y \sim \mathcal{N}(\mathbf{w}^T \mathbf{x}, \sigma^2)$

Recall that the pdf of a Gaussian random variable $\epsilon \sim \mathcal{N}(\mu, \sigma^2)$ is given by:

$$p(\epsilon \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\epsilon-\mu)^2}$$

Also, the pdf of a Laplacian random variable $z \sim \text{Laplace}(\mu, b)$ is given by:

$$p(z \mid \mu, b) = \frac{1}{2b} e^{-\frac{1}{b}(|z-\mu|)}$$

We are going to create our linear regression model with the assumption that each weight $w_j, j \in \{1 \dots M\}$ is drawn from $\text{Laplace}(0, b)$. Finally, assume we have a dataset of N i.i.d. samples $\mathcal{D} = \{(y^{(i)}, \mathbf{x}^{(i)})\}_{i=1}^N$

1. [3 pts] Derive the Laplace prior on the weights $p(\mathbf{w})$ given hyperparameter b . Write your answer in terms of the individual weights w_j , hyperparameter b , and dimension M .

$p(\mathbf{w})$

2. **[3 pts]** With your answer from the previous part, derive the conditional likelihood times the prior, $p(y^{(1)}, \dots, y^{(N)} \mid \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}, \mathbf{w}) p(\mathbf{w})$, which is proportional to the posterior $p(\mathbf{w} \mid \mathcal{D})$.

Note: Don't include any probability distribution shortcuts in your answer, e.g. make sure to write out the Gaussian distribution rather than stopping at an expression with $\mathcal{N}$ in it.

Your Answer

3. **[3 pts]** Derive the log of the conditional likelihood times the prior, $\log(p(y^{(1)}, \dots, y^{(N)} \mid \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}, \mathbf{w}) p(\mathbf{w}))$.

Your Answer

4. **[3 pts]** Finally, prove that $\operatorname{argmin}_{\mathbf{w}} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2 + \lambda \|\mathbf{w}\|_1$ is equivalent to finding $\mathbf{w}$ by minimizing the negative log of the posterior with a Laplace prior.

Specifically, write λ in terms of the log posterior hyperparameters, b and σ^2 .

Your Answer

λ

5 Collaboration Policy

After you have completed all other components of this assignment, report your answers to the following collaboration questions.

1. Did you receive any help whatsoever from anyone in solving this assignment? If so, include full details including names of people who helped you and the exact nature of help you received.

Your Answer

2. Did you give any help whatsoever to anyone in solving this assignment? If so, include full details including names of people you helped and the exact nature of help you offered.

Your Answer

3. Did you find or come across code that implements any part of this assignment? If so, include full details including the source of the code and how you used it in the assignment.

Your Answer