Assignment 1

Machine Learning I

Indian Statistical Institute

Instructor: Swagatam Das

Fall 2024

Total Points: 200

Notes

- Please complete and submit the assignments by the given deadline.
- The written assignments must be done in $\LaTeX$ unless otherwise specified.
- The programming assignments should be done in Python 3+.

1 Problems on vectors and matrices

1.1 Problem 1

[**Points:** $4 \times 3 = 12$] Let us consider the following matrix:

$$\begin{pmatrix} 1 & 1 & 2 \\ 3 & 5 & 7 \\ 0 & 2 & 1 \end{pmatrix} \tag{1}$$

1. What is the rank of this matrix?
2. What is the nullity of this matrix?
3. What is the row space of this matrix?
4. What is the null space of this matrix?

1.2 Problem 2

[**Points:** $4 \times 3 = 12$] Consider the systems of equations $Ax = B$.

1. When will we have an exact solution for such a system?
2. When will such a system have infinitely many solutions?
3. When will there be no solution for such a system?
4. What is meant by the least square solution of such systems?

1.3 Problem 3

[**Points:** $6 \times 2 = 12$] Let a function $f : \mathbb{R}^3 \to \mathbb{R}^3$ be as follows:

$$f(x_1, x_2, x_3) = (x_1^2,\ 2(1 - x_1)x_2,\ x_3^2). \tag{2}$$

1. Find the Jacobian matrix.
2. Find the Hessian matrix.

1.4 Problem 4

[**Points:** $6 \times 2 = 12$] Let a function $f : \mathbb{R}^3 \to \mathbb{R}$ be as follows:

$$f(x_1, x_2, x_3) = 2x_1(x_1 + 1) + x_2^2 + x_3^2 + x_2(2x_1 - x_3). \tag{3}$$

1. Find the optimum of f.
2. Can you comment on the nature of the optimum of f.

1.5 Problem 5

[**Points:** $6 \times 2 = 12$] Find the Eigen values and Eigen vectors of the following matrix:

$$\begin{pmatrix} 1 & -2 & 3 \\ -3 & 5 & 2 \\ 7 & -8 & 0 \end{pmatrix} \tag{4}$$

2 Problems on linear regression

2.1 Problem 1

[Points: 10] Let us take 3 points namely (2, 5), (4, 1) and (6, 7) lying on the two-dimensional plane. Find the line which is the closest to all these three points. Show all the steps of your derivation.

2.2 Problem 2

[Points: 30] Recall that the objective function for L_2-regularized linear regression is:

$$J(\mathbf{w}) = \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{w}\|_2^2, \tag{5}$$

where X is the design matrix (the rows of X are the data points).

- Prove that The global minimizer of J is given by:

$$\mathbf{w}^* = (X^T X + \lambda I)^{-1} X^T \mathbf{y}. \tag{6}$$

- Consider running the Newton's method to minimize J. Let $\mathbf{w_0}$ be an arbitrary initial guess for Newton's method. Show that $\mathbf{w_1}$, the value of the weights after one Newton step, is equal to $\mathbf{w}^*$.

2.3 Problem 3

[Points: 30] Let's take a look at a more complicated version of ridge regression called Tikhonov regularization. We use a regularization parameter similar to λ, but instead of a scalar, we use a real, square matrix $\Gamma \in R^{d\times d}$ called the Tikhonov matrix). Given a design matrix $X \in R^{n\times d}$ and a vector of labels $\mathbf{y} \in R^n$, our regression algorithm finds the weight vector $\mathbf{w}^* \in R^n$ that minimizes the cost function:

$$J(\mathbf{w}) = \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2 + \lambda\|\mathbf{\Gamma}\mathbf{w}\|_2^2. \tag{7}$$

- Derive the normal equations for this minimization problem—that is, a linear system of equations whose solution(s) is the optimal weight vector $\mathbf{w}^*$. Show your work. (If you prefer, you can write an explicit closed formula for $\mathbf{w}^*$)

- Give a simple, sufficient and necessary condition on Γ (involving only Γ not X nor y) that guarantees that $J(\mathbf{w})$ has only one unique minimum $\mathbf{w}^*$. (To be precise, the uniqueness guarantee must hold for all values of X and y, although the unique $\mathbf{w}^*$ will be different for different values of X and y.) (A sufficient but not necessary condition will receive part marks.)

Suppose you solve a Tikhonov regularization problem in a two-dimensional feature space ($d = 2$) and obtain a weight vector $\mathbf{w}^*$ that minimizes $J(\mathbf{w})$. The solution $\mathbf{w}^*$ lies on an isocontour of $\|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2$ and on an isocontour of $\|\mathbf{\Gamma}\mathbf{w}\|_2^2$ Draw a diagram that plausibly depicts both of these two isocontours, in a case where $\Gamma$is not diagonal and $y \neq 0$. (You don't need to choose specific values of X, y, or Γ your diagram just needs to look plausible.) Your diagram must contain the following elements:

- The two axes (coordinate system) of the space you are optimizing in, with both axes labeled.

- The specified isocontour of $\|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2$ labeled.

- The specified isocontour of $\|\mathbf{\Gamma}\mathbf{w}\|_2^2$ labeled.

- The point $\mathbf{w}^*$.

  These elements must be in a plausible geometric relationship to each other.

2.4 Problem 2: Coding exercise

[**Points:**$15 \times 4 = 60$] Let's say you want to buy a new flat and thus need to estimate a bank loan. You surveyed the newspapers for the price of 10 random flats spread across your city. However, the surveyed flats are of different sizes having diverse range of essential amenities. Thus, you created the following Table 1 listing the size of a flat, the number of bedrooms in that flat, and the corresponding price. You want to buy a flat which is about 950-1050 sq. ft. in size having either 2 or 3 bedrooms. Estimate the upper and lower limit of the bank loan given the data in Table 1.

Size of flat in sq. ft.	Number of bedrooms	Price in millions
1600	3	8.2
1260	2	6.6
1800	4	10.3
600	1	1.7
850	2	3.6
920	2	4.4
1090	2	5.4
890	2	4.8
1340	3	10.5
1650	2	7.4

Table 1: Table of flat sizes, number of bedrooms, and prices.

- **Task 2.2.a** Write a code which will provide you the least square estimation by solving the closed form solution for such problems.
- **Task 2.2.b** Write a code which will provide you the least square estimate by solving the problem using a gradient decent approach. Plot the convergence of the model parameters over successive iterations.
- **Task 2.2.c** Plot the flat prices as a function of flat size and number of bedrooms. Draw the least square estimators obtained respectively by solving the closed form solution and by gradient descent. Plot your least square estimations for the bank loan in both cases.
- **Task 2.2.d** Download the Portland House Price Prediction Dataset `https://www.kaggle.com/kennethjohn/housingprice`. Report the 10-fold cross validation mean squared error of your least square estimation model trained by gradient descent.