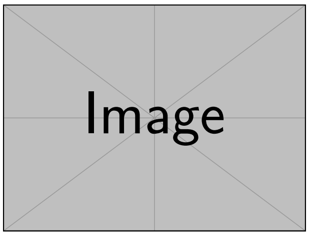
Mathematics Class Report

Assignment Title



Your Name
Course: MVA – Time Series Analysis

Instructor: Dr. ______
Deadline: ________

Academic Year: 2023–2024

Contents

1 Introduction

- Briefly describe the objective of your report.
- Provide context and motivation.

2 Theory

Definition

Here, you can state and define important mathematical concepts.

Theorem

Place for theorems, lemmas, properties, etc.

Question 1
Consider the following Lasso regression:

$$\min_{\beta \in \mathbb{R}^p} \frac{1}{2}\|y - X\beta\|_2^2 \quad + \quad \lambda\|\beta\|_1 \tag{1}$$

where $y \in \mathbb{R}^n$ is the response vector, $X \in \mathbb{R}^{n\times p}$ the design matrix, $\beta \in \mathbb{R}^p$ the vector of regressors and $\lambda > 0$ the smoothing parameter.

Show that there exists $\lambda_{\max}$ such that the minimizer of (1) is $\mathbf{0}_p$ (a p-dimensional vector of zeros) for any $\lambda > \lambda_{\max}$.

Answer 1

$$\lambda_{\max} = \max_{j\in\{1,\dots,p\}} \left|\sum_{i=1}^{n} X_{ij} y_i\right| \tag{2}$$

Proof: Let $f(\beta) = \frac{1}{2}\|y - X\beta\|_2^2$. The function is twice differentiable everywhere, and the Hessian is $\nabla^2 f(\beta) = X^T X \succeq 0$. Therefore f is convex. We know the function satisfies, for every pair of points (x, y): $f(y) \geq f(x) + \langle \boldsymbol{\nabla} f(x), y - x\rangle$, and so by letting $y = \beta$, $x = 0$, and computing the gradient of the function at 0 which is $\boldsymbol{\nabla} f(0) = -X^T y$ we finally get the identity:

$$f(\beta) \geq f(0) - \langle X^T y, \beta\rangle \tag{3}$$

Now, if we let $\lambda > \lambda_{\max}$, we have that $\lambda|\beta_i|$ is strictly greater than the absolute value of the i'th component of the dot product (this is true by definition of $\lambda_{\max}$), and therefore by adding all the terms we get

$$f(\beta) + \lambda\|\beta\|_1 > f(0) \quad \forall\beta \in \mathbb{R}^p$$

This proves what we wanted.

Question 2
For a univariate signal $\mathbf{x} \in \mathbb{R}^n$ with n samples, the convolutional dictionary learning task amounts to solving the following optimization problem:

$$\min_{(\mathbf{d}_k)_k,(\mathbf{z}_k)_k \|\mathbf{d}_k\|_2^2 \leq 1} \quad \left\| \mathbf{x} - \sum_{k=1}^{K} \mathbf{z}_k * \mathbf{d}_k \right\|_2^2 \quad + \quad \lambda \sum_{k=1}^{K} \|\mathbf{z}_k\|_1 \tag{4}$$

where $\mathbf{d}_k \in \mathbb{R}^L$ are the K dictionary atoms (patterns), $\mathbf{z}_k \in \mathbb{R}^{N-L+1}$ are activations signals, and $\lambda > 0$ is the smoothing parameter.

Show that

- for a fixed dictionary, the sparse coding problem is a lasso regression (explicit the response vector and the design matrix);
- for a fixed dictionary, there exists $\lambda_{\max}$ (which depends on the dictionary) such that the sparse codes are only 0 for any $\lambda > \lambda_{\max}$.

Answer 2

To proof the first statement, we will formulate the problem as a minimization over $z \in \mathbb{R}^{K(N-L+1)}$, where this vector is obtained from concatenating the K vectors $z_k \in \mathbb{R}^{N-L+1}$. We notice first that with this notation we have

$$\sum_{k=1}^{K} \|\mathbf{z}_k\|_1 = \|z\|_1$$

To put this problem in the lasso regression form, we notice that the convolution $z_k * d_k = M_{d_k} z_k$, where $M_{d_k} \in \mathbb{R}^{N \times (N-L+1)}$ is the matrix obtained by reversing d_k and putting one more coeficient at each row in a circular way.

$$M_{d_k} = \begin{bmatrix} (d_k)_L & 0 & \cdots & \cdots & 0 \\ (d_k)_{L-1} & (d_k)_L & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & (d_k)_1 \end{bmatrix}$$

Now, if we take the vector $y = \sqrt{2}x$ and we take the matrix $M \in \mathbb{R}^{N \times K(N-L+1)}$ as

$$M = \sqrt{2}\left[M_{d_1} \mid M_{d_2} \mid \cdots \mid M_{d_K} \right]$$

We have that the problem is indeed a lassso regression minimization one.

The second statement is an immediate consequence of the first question, applied to the vector y and the matrix M.

3 Methodology

- Explain the approaches, algorithms, or tools used.
- Illustrate with formulas:

$$x(t) = \sum_{i=1}^{n} a_i \cdot \phi_i(t)$$

4 Results

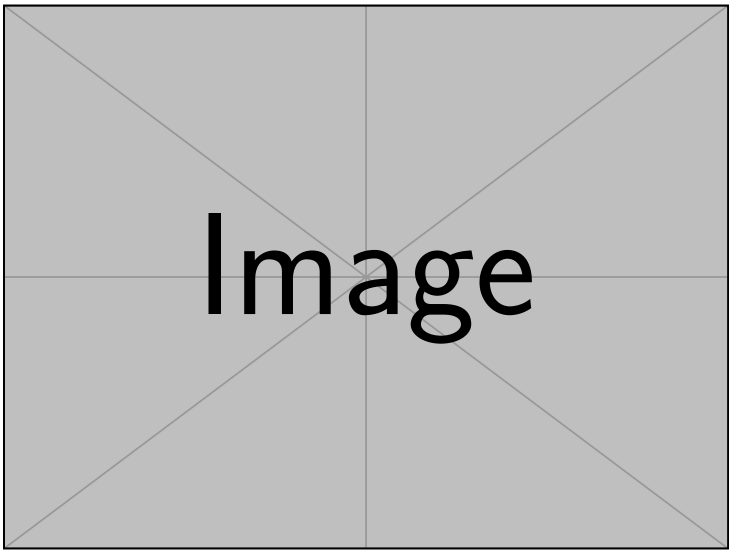


Figure 1: Sample Plot – describe your results here.

Parameter	Value	Description
a	2	Slope coefficient
b	3	Intercept

Table 1: An example table

5 Discussion

- Interpret and critique your results.
- Discuss any surprising findings.

6 Conclusion

- Summarize what you learned.
- Propose future directions or applications.

References

1. Author, "Title of Book or Article," Journal/Publisher, Year.
2. ...