INFO371 Lab 7: Regularization

Your name:

Deadline: Wed, Feb 21, 11:59pm

Introduction

Please submit the completed lab by end of the day. You should submit a) your code (notebooks, rmd, whatever) and b) the lab in a final output form (html or pdf).

Please do not just provide computer output. Always comment on your main findings. Include any substantial comments as a separate text blocks. Also limit your output: do not submit pages and pages of whatever your code spits out.

Note: you may want to do some of it on paper instead of computer. You are welcome to do this but please include the result as an image into your final file.

Working together is fun and useful but you have to submit your own work. Discussing the solutions and problems with your classmates is all right but do not copy-paste their solution! Please list all your collaborators below:

1.

2. ...

Loss Function and Regularization

Your task is to explore the loss function of linear regression, the regularization penalty, and the penalized loss function.

Data

Here we use a simple artificial dataset:

x	y
1.00	1.00
3.00	6.00
6.00	7.00
8.00	5.00

1 Linear Regression

1.1 Plot the data

Create a simple plot to understand the shape of the data.

1.2 Linear Regression

Estimate two regression models:

$$y_i = \beta^1 x_i + \epsilon_i \quad \text{and} \quad y_i = \beta^1 x_i + \beta^2 x_i^2 + \epsilon_i.$$

Note: this regression does *not* include the constant! We do not include the constant to reduce the dimensionality of the problem: now it is just 1 and 2-dimensional, instead of being 2 and 3-dimensional.

Report and comment the regression coefficients and t-values. Comment also the regression lines: which one do you think is more reasonable to pick?

2 Loss function

2.1 Plot the loss function

Here your task is to contour-plot the loss function for the second linear regression model $y_i = \beta^1 x_i + \beta^2 x_i^2 + \epsilon_i$.

1. Create the grid. The predicted value $\hat{y}_i$ includes two parameters: β^1 and β^2. Create a grid (say, 100×100) of both of these parameters. The grid should include the point (0,0) and the estimated values (β^1, β^2).
2. Compute the loss function (sum of squared errors) at each point on this grid.
3. Display the loss function with contourplot. Mark the location of optimum (β^1, β^2), and $\beta^1 = 0$ and $\beta^2 = 0$ lines.

3 L1 regularization

Now let's analyze lasso (L_1) regularization.

3.1 Regularization penalty

L_1 regularization penalty is computed as $P(\beta) = \sum_{i=1}^{K} |\beta^i|$.[1] Note: the intercept β^0 is not included. As we don't have intercept in our model, it does not matter much.

- Take the same grid you were using above. Compute the the penalty function at each gridpoint.
- plot the penalty function using a similar contourplot as you did above. Mark the location of optimum (β^1, β^2), and $\beta^1 = 0$ and $\beta^2 = 0$ lines.
- Where is the smallest penalty value on this figure?

[1] L_1 is a form of general L_p metric: $l_p = (\sum_i |\beta^i|^p)^{1/p}$. It is also called "Manhattan metric" or "taxicab metric".

4 Lasso regression

Finally, lets put these two pictures together. The lasso regression minimizes the sum of quared deviations + L_1 penalty parameters:

$$\min_{\beta}(y - X\beta)'(y - X\beta) + \lambda \sum_{i=1}^{K} |\beta^i| \tag{0.1}$$

where λ is the amount of penalty we are introducing to the model (regularization parameter).

1. Choose a λ value.
2. Create the penalized loss function by adding the sum of squared deviations and the penalty, as in (0.1).
3. Produce a few images with different λ values in a way that the optimum get close to the original, unregularized optimum, and close to zero. Comment your findings.