Pattern Recognition and Machine Learning

Home Work 4: Logistic Regression and Regularized Regression

Faculty of New Science & Technologies
University Of Tehran
Fall 2019

Part 1

General Homework Policies

1. Due date of this homework is on *Saturday 23 Azar 98 (14 Dec 2019)*, so you need to submit it before the due date[midnight 23 Azar] otherwise you won't get any score!
2. Try to budget your time because due dates are hardly changeable, and we will not accept late homework for any reason.
3. You are welcome to collaborate, cooperate, and consult with your classmates provided that you write-up the solutions independently.
4. Don't plagiarize! Write everything in your own words, and properly cite every outside source you use. Taking credit for work as well as ideas that are not your own is plagiarism. Students who plagiarize will not get any score and they will be introduced in the class.
5. Please create reference for all sources(books, papers, websites) which you use.
6. Please create a cover letter for your report which is simply is the Homework#, title of the course, your name, surname, and student number.
7. You may post questions asking for clarifications and alternate perspectives on concepts on piazza or in the class.
8. Email your final file of assignment to *sajjadaghapour@ut.ac.ir*, *ah.havvaei@ut.ac.ir* and *e.sadeqi.n@gmail.com* with subject [PRML hw# Surname] which # indicates number of the home work.

Part 2

Questions

1 Problems

1.1 Ridge and Lasso

(i) Given $y \in \mathbb{R}^n$, consider ridge regression with predictor matrix $X = I_{n\times n}$, i.e.,

$$\hat{\beta}^{ridge} = \arg\min_{\beta\in\mathbb{R}^n} \|y-\beta\|_2^2 + \lambda\|\beta\|_2^2 \tag{2.1}$$

$$= \arg\min_{\beta\in\mathbb{R}^n} \sum_{i=1}^{n}(y_i-\beta_i)^2 + \lambda\sum_{i=1}^{n}\beta_i^2 \tag{2.2}$$

Show that the solution is

$$\hat{\beta}_i^{ridge} = \frac{y_i}{1+\lambda},\ i=1,\dots,n \tag{2.3}$$

(ii) For the lasso with identity predictor matrix,

$$\hat{\beta}^{lasso} = \arg\min_{\beta\in\mathbb{R}^n} \|y-\beta\|_2^2 + \lambda\|\beta\|_1 \tag{2.4}$$

$$= \arg\min_{\beta\in\mathbb{R}^n} \sum_{i=1}^{n}(y_i-\beta_i)^2 + \lambda\sum_{i=1}^{n}|\beta_i| \tag{2.5}$$

the solution is

$$\hat{\beta}_i^{lasso} = \begin{cases} y_i+\lambda/2 & y_i < -\lambda/2 \\ 0 & |y_i| \le \lambda/2 \\ y_i-\lambda/2 & y_i > \lambda/2 \end{cases},\ i=1,\dots,n \tag{2.6}$$

For a fixed value of λ (e.g., you can take $\lambda = 1$), draw $\hat{\beta}_i^{ridge}(y_i)$ and $\hat{\beta}_i^{lasso}(y_i)$ as functions of y_i. Describe the difference between these two coefficient functions.

1.2 Logistic regression

Solve Exercises 8.6 and 8.7 of Murphy's Textbook.

2 Programming: Spam classification using logistic regression

See exercise 8.1 of Murphy for the description of the dataset and the task. You should provide the following tasks,

- Perform logistic regression based on gradient descent algorithm.
- Report your error rate and draw the cost function for each iteration of your functions.
- Compare the achieved results with the naive Bayes classifier that you previously done in the homeworks.

Good Luck!