
Machine Learning

Problem Set 4

Hesam Montazeri
Fereshteh Fallah
Mozhgan Mozaffari Legha
Farvardin 8, 1399
(March 27, 2020)

Problem 1: Review part

Write your reviews for the whiteboard notes and the slides of the lectures of this week. Write down all formulas and explain in detail each step of the derivations, if applicable.

Problem 2: Conceptual questions and math review

- (a) [ISL] chapter 4: questions 1, 4, 6, 7, 8, 9.
- (b) Derive the second order Taylor approximation of $f(x, y) = 3x^2 - 2xy + y^3$ at point $[0, 1]^T$.
- (c) Using the chain rule, calculate $\partial z / \partial u$ and $\partial z / \partial v$ using the following functions:
$$z = f(x, y) = x^2 + xy - 4y^2, \quad x = x(u, v) = 5u - v, \quad y = y(u, v) = u + 3v$$

Problem 3: Exponential family distribution

Write down the negative binomial and beta distributions as exponential family distributions.

Problem 4: Newton's method for linear regression

Derive the updating formula of Newton's method for linear regression.

Problem 5: Programming: logistic regression

Implement the parameter estimation algorithm for the logistic regression using iteratively reweighted least squares method. You may use the existing implementation for weighted linear regression. Show the correctness of your code by comparing to the built-in functions in R.

Problem 6: Programming: high-dimensional classification

Your task is to build a classification model for the uploaded simulated data (HW4train.RData) with 150 observations and 1000 features.

- (a) Logistic regression
- (b) Logistic LASSO regression
- (c) Logistic ridge regression
- (d) First use forward stepwise selection and then perform logistic regression (you may use *step* function in R).

Report the generalization error of each model. Write the predictions of each model for the uploaded test data (HW4test.RData) in a csv file. You may use built-in functions in R. Extra credit for the most accurate generalization error and predictions.

We encourage discussing the problems with other students, however, similarity between solutions is not allowed. (**Important**) Studying any online or previous solutions, no matter to what extent, is strictly forbidden and is considered as a violation of the academic honor code. Submit your solutions by Farvardin 16, 1399.