Due: Nov 15, 2021, 12:00 PM

Problem 1 - Linear Regression

0.1

Let $y = w_0 + w_1 x + \epsilon$, where ϵ is a normally distributed random error with mean 0 and variance σ^2. Consider fitting a least squares regression model $\hat{y} = \hat{w_0} x + \hat{w_1}$.

- In linear regression, why do we usually minimize the square error ($\sum (y - \hat{y})^2$) rather than l_1-norm ($\sum |y - \hat{y}|$)?
- Denote the residuals as $e = y - \hat{y}$. Are the residuals e and fitted values $\hat{y}$ correlated? Why or why not? [Hint: Simple linear regression is a special case of multivariate regression, thus conclusion of multivariate regression could still be applied, such as $\hat{Y} = HY$, where $H = X(X^T X)^{-1} X^T$.]

0.2

Let

$$Y_1 = 2w_1 + w_2 + \epsilon_1$$
$$Y_2 = w_1 - w_2 + \epsilon_2$$
$$Y_3 = 3w_1 + \epsilon_3$$

where $(\epsilon_1, \epsilon_2, \epsilon_3)$ are independent normal errors with mean 0 and variance σ^2.

- Find the least squares estimator of (w_1, w_2), express it as a 2-dimensional vector $(\hat{w_1}, \hat{w_2})$ where each element is a function of the labels Y_1, Y_2, Y_3.
- Compute the covariance matrix of $(\hat{w_1}, \hat{w_2})$. (Note that your answer should not include the labels Y_1, Y_2, Y_3.)

Problem 2 - Logistic Regression

- MSE (mean squared error) is often used as the loss function in regression models. Why can't we use it in logistic regression?
- Logistic function ranges from 0 to 1. Explain how parameter β controls the response variable. For example $g(p) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, where $g(p) = ln \frac{p}{1-p}$, what is the influence of coefficients β_i.

Problem 3 - Problem 3 (Programming) - Modern Regression

Read about and download the blogFeedback dataset from here:
https://archive.ics.uci.edu/ml/datasets/BlogFeedback

1. Perform Least Squares Regression, Ridge Regression, and LASSO to predict the target variable. You can use any package to do this, but ensure that the parameters are crossvalidated. Train on **blogData_Train.csv** and test on **blogData_test-2012.03.31.01_00.csv**. Report RMSE for each model.
2. What are the most important features according to LASSO?