CS436/CS580L Homework 2

Fall 2018, Xi Peng

Instruction:

(a) Release: Oct 25, 2018.
(b) Due: Nov 6, 2018.
(c) 10 points for each question.

You need to upload a report to answer all the questions. The source codes should be directly pasted into the report as the appendix.

Question 1.

In this problem you will replicate and expand upon the prototype data analysis problem in lecture slides "L5&6_Linear Regression_Gradiant Descent_Polynomial Regression 2.pdf". In this case someone has generated data by sampling the sinusoidal function $f(x) = \sin(2\pi x), x \in [0, 1]$. However, pretend not to know this fact and try to fit different models to discover the relationship between x and y.

1. Generate a dataset $\mathcal{D}$ consisting of $N = 200$ by evaluating the function $f(x)$ at N uniformly spaced points in x and adding iid white Gaussian noise with deviation $\sigma = 0.1$. I.e., $x_i = \frac{i-1}{N-1}$ *and* $y_i = x_i + \epsilon_i, \epsilon_i \sim \mathcal{N}(0, \sigma^2), i = 1, \cdots, N,$ *and* $\mathcal{D}_N = \{(x_{i,} y_i)\}_{i=1}^{N}$.
2. Repeat the polynomial fitting experiment by selecting a random subset of $\mathcal{D}$ of the appropriate size to train the model. Report estimate coefficients (model parameters) ω, MSEs, both as a function of the chosen training set size and the regularization λ.
3. Select a test dataset $\mathcal{D}_{test}$ as $N_{test} = 199$ uniformly spaced points $x_{i,test} = x_{i,train} + \frac{1}{2\cdot(N_{train}-1)}, i = 1, \cdots, N_{train} - 1, y_i = f(x_i)$, and plot the comparison of training and test MSEs for the fitted models as a function of N_{train} and λ. Choose at lease 10 different values of both parameters for these graphs.
4. What is the conditional correlation of two arbitrary points $y_i|x_i$ and $y_j|x_j$, i.e., $p(y_i, y_j|x_i, x_j)$?
5. Support that any two samples of $f(x)$ have the covariance function $\sigma(y_i, y_j) = 0.1 \cdot \exp\{-\frac{(x_i-x_j)^2}{2\cdot 0.01}\}$. Suppose further that the joint distribution of any N such samples, conditioned on x_i's, is Gaussian with mean zero and the covariance matrix given by the $\sigma(y_i, y_j)$ function. What is the conditional distribution of $y(x)$ at some new arbitrary point x, given a dataset of N_{train} training points? In other words, what is $P(y|x, \mathcal{D}_{train})$?
6. Use the conditional model found in the previous step in place of the standard curve fit model. Since the conditional model is defined in terms of a distribution, pick the mean prediction, $E(y|x, \mathcal{D}_{train})$ to compare with the standard curve fitting model. Report this model's MSE on the training and testing sets, as a function of N.
7. Contrast performance of the two models. Which one has a better performance and why?

Question 2.

In this problem you will implement gradient descent algorithm from scratch. Use MATLAB build-in dataset for binary classification problem by running "load ionosphere". Use the first 301 data points as the training set and the following 50 data points as the testing set. Suppose we choose logistic regression as the classification model.

1. Implement gradient descent algorithm without regularization.
2. Implement stochastic gradient descent algorithm without regularization.
3. Implement gradient descent algorithm WITH regularization.

Notice:

a) Use functions as demonstrated in the class, e.g., LR_GD(), LR_SGD(), LR_Test(), Sigmoid(), so some functions can be shared in the three questions.
b) For all the three questions, you need to try at least 3 different learning rates, or design the learning rate as a function of iterations. Plot the training and testing errors with respect of iterations. What do you observe by using different learning rate? What is the best learning rate? Explain how to find it?
c) For the third question, you also need to try at least 3 different weights for the regularization term. Plot the training and testing errors with respect of iterations. What do you observe by using different weights? What is the best weight? Explain how to find it?