COL 341: Assignment 1

Notes:

- This assignment has two parts - Linear Regression and Logistic regression.
- You should submit all your code (including any pre-processing scripts written by you) and any graphs that you might plot.
- You are advised to use vector operations (wherever possible) for best performance.
- Include a report of maximum 3 pages which should be a brief description for second question explaining what you did. Include any observations and/or plots required by the question in the report.
- You should use Python/R for all your programming solutions.
- Your assignments will be auto-graded, make sure you test your programs before submitting. We will use your code to train the model on training data and predict on test set. You will need to perform cross validation to figure out regularization parameters during auto-grading as well.
- Input/output format, submission format and other details are included on 'README.md'. Your programs should be modular enough to accept specified parameters.
- You should submit work of your own. You should cite the source, if you choose to use any external resource. You will be awarded F grade or DISCO in case of plagiarism.
- You can use total of 7 buffer days across all assignments.
- You can download data from this link.

1. **Linear Regression (50 points, Release date: Aug. 3, 2018, Due date: Aug. 10, 2018)**
   In this problem, we will use Linear Regression to predict release year of a song from audio features. You have been provided with train and test split of the dataset. Data is given in csv format with target as first value, please refer to README file for more details and submission format. For the details of the original dataset, you are encouraged to look at this webpage.

   (a) (12.5 points) Given a training dataset $D = \{(x^{(i)}, t^{(i)})\}_{i=1}^{m}$, recall that linear regression optimization problem can be written as:

   $$\min_{w,b} \frac{1}{2m} \sum_{i=1}^{m} ||w^T x^{(i)} + b - t^{(i)}||^2 \tag{1}$$

   Here, (w, b) represents the hyper-plane fitting the data. Implement linear regression on this dataset using normal equations (analytic solution).

   (b) (12.5 points) Implement ridge regression using normal equations to find the optimal hyper-plane. You should use 10 fold cross-validation to determine optimal value of regularization parameter. Please don't shuffle the data during cross-validation.

   (c) (25 points) Feature creation and selection [1] [2] are important part of machine learning. Use Lasso to learn hyper-plane for the given training data. You are free to use 'Lasso model fit with Least Angle Regression (Lars)' package/function for this part. You should try out different transformations to get best performance.
   Note: LARS package is available for R and python.

Evaluation:

- Normalized mean squared error will be used as evaluation criterion.
- For part-a and part-b, you can get 0 (error), half (code runs fine but predictions are incorrect within some predefined threshold) and full (works as expected).
- For part-c, marks will be given based on error on test dataset. There will be relative marking for this part.

2. **Logistic Regression (50 points, Release date: Aug. 3, 2018, Due date: Aug. 17, 2018))**
In this problem, we will use Logistic Regression to build a binary text classifier for 'imdb' dataset. We will be solving the logistic regression optimization problem using the gradient descent algorithm. You must use given vocabulary for uniformity and word counts as features. Do not shuffle data within your code, input data will already be shuffled. For details of the original dataset, you are encouraged to look at this webpage.

   (a) **(30 points)** Given a training dataset $D = \{(x^{(i)}, t^{(i)})\}_{i=1}^{m}$, recall that the log-likelihood for logistic regression can be written as:

$$L(\theta) = \sum_{i=1}^{m} t^{(i)} \log(h_\theta(x^{(i)})) + (1 - t^{(i)}) \log(1 - h_\theta(x^{(i)})) + \frac{\lambda}{2}||w||^2 \qquad (2)$$

   Here, $h_\theta(x) = \frac{1}{1+e^{-w^T x + b}}$, (w, b) represents the decision surface learned by logistic regression. Implement gradient descent algorithm and solve the logistic regression problem using it. Learning rate is a critical element in gradient descent based algorithms, you should experiment with following learning rate strategies: i) Constant learning rate, ii) Adaptive learning rate as $\eta_t = \frac{\eta_0}{\sqrt{t}}$, iii) Adaptive learning rate using exact line search algorithm (using binary search). Here, η_0 is the initial learning rate and t is the time step i.e. iteration. Use 10-fold cross validation to determine optimal value of regularization parameter. You should plot value of log-likelihood against number of floating point operations i.e. basic mathematical operations such as addition, subtraction, multiplication and division for each case. Include your observations in your report.

   (b) **(15 points)** Implement stochastic gradient descent algorithm and solve the logistic regression problem using it. Use batch size of 128. Experiment with different learning rate strategies: i) Constant learning rate, ii) Adaptive learning rate as $\eta_t = \frac{\eta_0}{\sqrt{t}}$, iii) Adaptive learning rate using exact line search algorithm (using binary search). Use 10-fold cross validation to determine optimal value of regularization parameter. You should plot value of log-likelihood against number of floating point operations for each case. Include your observations in your report.

   (c) **(5 points)** Stopping criterion is an important aspect of gradient descent. Experiment with any two stopping criteria of your choice and report your observations. Based on your experiments in part-a & b, which learning rate strategy leads to best convergence? You are encouraged to include graphs to support your conclusion.

Evaluation:

- Marking will be done in two parts: code (75%) and report(25%).
- For code: you can get 0 (error), half (code runs fine but predictions are incorrect within some predefined threshold) and full (works as expected).

Extra Readings:

1. Regression Shrinkage and Selection Via the Lasso
2. Compressive Sensing Resources
3. Least Angle Regression