ELL409
Assignment 2: Report

-Gaurav Agrawal
2017ME20664

1. Multinomial Logistic Regression

Database:

MNIST database, downloaded from Kaggle given for Assignment 1 of the course.
The train file consist of first column as the Labels of the digit. The 28x28 image of digits has been flattened into an array of 1x784 and appended next to the label, making the first entry of 1x785. The test only consists of flattened pixels and no labels. So the dimension of row will be 1x784 corresponding to 28x28 pixels of some digit.

Procedure:

- Preprocessing of data: 1st row is labels, stored them in Y and converted it into one-hot representation. Followed by 784 rows for X (iput).
- Activation function: Defined the softmax function for multi-class classification.
- Loss function: Used cross-entropy loss for multiclass.
- Calculated gradient, for updating weights. After Mathematics it turns out to be gradient = (input*error).
- Initialise weights using the random number generator from gaussian distribution using numpy
- Performed iterations uodating the weights using gradient descent at each step.

Result:

Classifies each image as the corresponding digit on the test data set.

Difference:

1. Linear regression requires to establish the linear relationship among dependent and independent variable whereas it is not necessary for logistic regression.
2. In the linear regression, the independent variable can be correlated with each other. On the contrary, in the logistic regression, the variable must not be correlated with each other.

Advantages:

1. It is more robust to violations of assumptions of multivariate normality and equal variance-covariance matrices across groups.
2. It is similar to linear regression, but more easily interpretable diagnostic statistics.
3. MLR does not assume a linear relationship between the dependent and independent variables.
4. MLR does not require that the independents be unbounded.
5. Normally distributed error terms are not assumed.

Disadvantages:

1. It cannot predict continuous outcomes. For example, logistic regression could not be used to determine how high an influenza patient's fever will rise, because the scale of measurement -- temperature -- is continuous.
2. It requires that each data point be independent of all other data points. If observations are related to one another, then the model will tend to overweight the significance of those observations. This is a major disadvantage, because a lot of scientific and social-scientific research relies on research techniques involving multiple observations of the same individuals. For example, drug trials often use matched pair designs that compare two similar individuals, one taking a drug and the other taking a placebo. Logistic regression is not an appropriate technique for studies using this design.

2. Bayesian Linear Regression

Database:

The modified Boston housing dataset consists of 489 data points, with each datapoint having 3 features. This dataset is a modified version of the Boston Housing dataset found on the UCI Machine Learning Repository.

Features

- RM: average number of rooms per dwelling
- LSTAT: percentage of population considered lower status
- PTRATIO: pupil-teacher ratio by town

Target Variable MEDV: median value of owner-occupied homes

Procedure:

- Initialised X as the first three rows (as explained above data has three features) and Y as the fourth row.
- Assumed values of hyperparameters: alpha and beta.
- Calculated optimal value of weight, using the formula of Bayesian linear regression.
- Predicted loss value of the output.

Result:

Model to predict the median value of the owner occupied homes on unknown set of features.

Difference:

In the Bayesian viewpoint, we formulate linear regression using probability distributions rather than point estimates. The response, y, is not estimated as a single value, but is assumed to be drawn from a probability distribution.

Advantages:

1. In general, the advantage of Bayesian estimation is that you can incorporate the use of a prior, or assumed knowledge about the current state of "beliefs", and how the evidence might update those beliefs.
2. By this Bayesian processing, you recover the whole range of inferential solutions, rather than a point estimate and a confidence interval as in classical regression.

Disadvantages:

1. It's computationally intensive. Especially for models involving many variables. For a large dataset with many variables being estimated, it may very well be prohibitively computationally intensive, especially in certain circumstances where the data cannot readily be thrown onto a cluster or the like. Some of the ways to resolve this, like augmented data rather than MCMC, are somewhat theoretically challenging, at least to me.
2. Posterior distributions are somewhat more difficult to incorporate into a meta-analysis, unless a frequentist, parametric description of the distribution has been provided.
3. Depending on what journal the analysis is intended for, either the use of Bayes generally, or your choice of priors, gives your paper slightly more points where a reviewer can dig into it. Some of these are reasonable reviewer objections, but some just stem from the nature of Bayes and how familiar people in some fields are with it.