CS5542 Big Data Analytics and Apps

Problem Set-3 (PS-3)

Deadline: March 20th 2018

1. Logistic Regression
2. Why use logistic regression rather than ordinary linear regression?

Linear Regression is used to establish a relationship between Dependent and Independent variables, which is useful in estimating the resultant dependent variable in case independent variable change. Logistic Regression on the other hand is used to ascertain the probability of an event. And this event is captured in binary format, i.e. 0 or 1.

1. What is the logistic curve?

An S-shaped curve that represents an exponential function and is used in mathematical models of growth processes

1. How are probabilities, odds and logits related?

They are different ways of stating the same information, like Fahrenheit and centigrade temperatures.Probabilities are stated as numbers between zero and one, and the sum the probabilities of all possible outcomes is one.Odds are stated in arbitrary scale, often chosen to use small integers. It's the chance of an event not happening and the chance of the event happening. So if the probability is p, the odds can be stated (1–p):p. But as I said it's usually scaled up to use small integers. If the probability is 0.5, you could write odds of 0.5:0.5, but no one ever does, they write 1:1. If p is 0.25, they write 3:1. If it's 0.2222, they write 7:2. Logit goes from negative infinity to infinity, it's the natural logatithm of p/(1–p). Logit is handy for some mathematical reasons.

1. What is an odds ratio?

Logistic regression is one way to generalize the odds ratio beyond two binary variables. Suppose we have a binary response variable Y and a binary predictor variable X, and in addition we have other predictor variables Z1, ..., Zp that may or may not be binary. If we use multiple logistic regression to regress Y on X, Z1, ..., Zp, then the estimated coefficient {\displaystyle {\hat {\beta }}\_{x}} {\hat {\beta }}\_{x} for X is related to a conditional odds ratio.

1. What is a loss function?

It describes how far off the result your network produced is from the expected result - it indicates the magnitude of error your model made on its prediciton. You can then take that error and 'backpropagate' it through your model, adjusting its weights and making it get closer to the truth the next time around.

1. What is a maximum likelihood estimate?

It is a method in statistics for estimating parameter(s) of a model for given data. The basic intuition behind MLE is that the estimate which explains the data best, will be the best estimator.The main advantage of MLE is that it has asymptotic property. It means that when the size of the data increases, the estimate converges faster towards the population parameter. We use MLE for many techniques in statistics to estimate parameters. I have explained the general steps we follow to find an estimate for the parameter.

1. Softmax Regression Model
2. What is the difference between logic regression and Softmax regression?

Logistic regression as a binary classifier and softmax regression is one way(there are other ways) to implement an multi-class classifier. The number of output layers in softmax regression is equal to the number of class you want to predict.

1. What are the limitations of Softmax regression?

It doesn’t provide null rejection. You need to specify cases to handle null values/labels while training the classifier. SoftMax regression won't work on data which is linearly separable.

1. What is the Derivative of Softmax? Why is it needed?

Due to the desirable property of softmax function outputting a probability distribution, we use it as the final layer in neural networks. For this we need to calculate the derivative or gradient and pass it back to the previous layer during backpropagation.

1. What is gradient of the cross entropy loss function? Why is it needed?

Technically no because "softmax loss" isn't really a correct term, and "cross-entropy loss" is. So cross-entropy loss is really the correct term to use. The softmax classifier is a linear classifier that uses the cross-entropy loss function. In other words, the gradient of the above function tells a softmax classifier how exactly to update its weights using something like gradient descent. So in short, they aren't the same. However, people use the term "softmax loss" when referring to "cross-entropy loss" and because you know what they mean, there's no reason to annoyingly correct them. Because they are used interchangeably, the two terms are effectively the same.

1. How is the regularization term used in Softmax regression?

Regularization does NOT improve the performance on the data set that the algorithm used to learn the model parameters (feature weights). However, it can improve the generalization performance, i.e., the performance on new, unseen data, which is exactly what we want.In intuitive terms, we can think of regularization as a penalty against complexity. Increasing the regularization strength penalizes "large" weight coefficients -- our goal is to prevent that our model picks up "peculiarities," "noise," or "imagines a pattern where there is none."Again, we don't want the model to memorize the training dataset, we want a model that generalizes well to new, unseen data.In more specific terms, we can think of regularization as adding (or increasing the) bias if our model suffers from (high) variance (i.e., it overfits the training data). On the other hand, too much bias will result in underfitting (a characteristic indicator of high bias is that the model shows a "bad" performance for both the training and test dataset). We know that our goal in an unregularized model is to minimize the cost function, i.e., we want to find the feature weights that correspond to the global cost minimum (remember that the logistic cost function is convex).

1. Softmax Regression: Case study

Solve the following questions by referring to the website http://www.pyimagesearch.com/2016/09/12/softmax-classifiers-explained/

1. Demonstrate how the calculations of the Softmax classifier/cross-entropy loss function and the weight matrix W.

It’s much easier for us as humans to interpret probabilities rather than margin scores (such as in hinge loss and squared hinge loss).

Furthermore, for datasets such as ImageNet, we often look at the rank-5 accuracy of Convolutional Neural Networks (where we check to see if the ground-truth label is in the top-5 predicted labels returned by a network for a given input image).

Seeing (1) if the true class label exists in the top-5 predictions and (2) the probability associated with the predicted label is a nice property.

1. Show how to perform gradient decent and other optimization algorithms.



1. Explain how to perform optimization using the regularization term to increase classification accuracy.

Notice that their classifier has obtained 65% accuracy, an increase from the 64% accuracy when utilizing a Linear SVM in our linear classification post.

1. Explain the cost function and classification accuracy for Softmax classifier for the image as airplane.

For each of the randomly sampled data points, we are given the class label probability for both “dog” and “cat”, along with the actual ground-truth label.

Based on this sample, we can see that we obtained 4 / 5 = 80% accuracy.

But more importantly, notice how there is a particularly large gap in between class label probabilities. If our Softmax classifier predicts “dog”, then the probability associated with “dog” will be high. And conversely, the class label probability associated with “cat” will be low.

Similarly, if our Softmax classifier predicts “cat”, then the probability associated with “cat” will be high, while the probability for “dog” will be “low”.

This behavior implies that there some actual confidence in our predictions and that our algorithm is actually learning from the dataset.

Exactly how the learning takes place involves updating our weight matrix W, which boils down to being an optimization problem. We’ll be reviewing how to perform gradient decent and other optimization algorithms in future blog posts.

The data:

