Project Report

MATH F424

Applied Stochastic Process

"Churn Analysis"

By

Janhavi Prabhu :- 2018B4PS0784G

Rajvi Sampat :- 2018B4A70820G

Padmashri Saravanan :- 2018B4PS1039G

Introduction

A Markov Chain is a model that tell us something about the probabilities of a sequence of random variables, or *states*. The property of a Markov Chain is that the states preceding the current state have no effect on the future states. The main elements of a Markov Chain are:

i. $Q = q_1,q_2,...,q_n$: A set of n states
ii. $A = a_{11},a_{12},....a_{nn}$: The transition probability matrix representing the probability a_{ij} of moving fromt state i to j.
iii. $P = p_1,p_2,...p_n$: An initial probability distribution

A Hidden Markov Model (HMM) on the other hand, consists of a set of hidden states and a set of observable events. In many cases, the events we are interested in, are hidden. A very simple example is predicting the weather based on the number of ice-creams a child has eaten. In addition to the terms we discussed in Markov Models, HMMs also consist of:

i. $b_i(o_t)$: A set of probabilities of observations, also called emission probabilties
ii. $O = o_1,o_2,...o_n$: The set of n observations

Churn Analysis

Churn Analysis is the process of predicting whether a customer is going to stop using a service, based on their session data, ie. How long they're logging in and how frequently. Numerous applications and services such as mobile apps, Netflix, etc, use the concept of Churn Analysis to predict churning customers. These customers are then incentivised to stay using rewards and discounts, since the cost of keeping a customer is much less than the revenue required to bring an entirely new customer.

We used to churn analysis on the Telco Dataset to predict churning customers.

Forward Algorithm

The forward algorithm is used to compute the likelihood of the observation sequences. Each 'cell' of the forward algorithm is computed by calculating all possible hidden state sequences before it. However, the reason for the efficiency of this algorithm is that it stores the value of the (k-1)th cell (the first k-1 observations) and uses it to calculate the likelihood of the kth observation. Essentially, each cell uses the following formula:

$$\alpha_t(j) = P(o_1,o_2,...o_t,q_t=j|\lambda)$$

For a given state q_j, $\alpha_t(j)$ is computed as follows:

$$\alpha_t(j) = \Sigma\alpha_{t-1}(j)a_{ij}b_j(o_t)$$

Viterbi Algorithm

The Viterbi Algorithm is used to figure out which sequence of hidden states is most likely from the given observation sequence. The Viterbi Algorithm also uses a concept similar to the Forward Algorithm, but here, we try to find the sequence of hidden states.
Each cell of the Viterbi Algorithm calculates:

$$v_t(j) = \max P(q_1,q_2,...q_{t-1},o_1,o_2,...o_t,q_t=j| \lambda)$$

The value of each cell is computed recursively by taking the maximum of probabilties which will give us the most probable path of reaching this cell.
The difference between the Viterbi and the Forward algorithm is that the Viterbi takes the maximum whereas the Forward takes the sum.

Baum-Welch Algorithm

The Baum-Welch Algorithm is used to train both the emission and transition probabilities. It's an iterative algorithm which starts with an estimate, and then improves on the estimate, learning better probabilities as it goes.

Telco Dataset Churn Analysis

We have made use of the Telco Dataset, which contains personal data of around 7000 customers. It contains information such as: Gender, Tenure, Monthly Charges, Payment Method and so on. Since this dataset is not a time series dataset, which is the kind of dataset usually used in Churn Analysis, we have made a few changes. Our hidden state is whether or not the customer is going to churn.
In order to choose our observations, we found 5 of the most correlated observations to the Churn column of the dataset. These 4 observations were: Monthly Charges, Paperless Billing, Payment Method and Senior Citizen. Using conditional probability, we found the probability of moving from state i to j given the previous states. Our final probability was the probability of churning, given the particular path of states. If the probability was beyond a certain threshold value, we say that the customer churns, else, the customer does not churn.

We divided the data into training and test set according to the ratio 70:30. Using the training set, we computed the probabilities of churning for each customer. Based on our results, we chose a threshold value and tested it on the test set.

Results

We tried different threshold values for the probability of churning, and finally got a precision of 0.76 for the threshold score of 0.5 for the training set. Using the same threshold score on the test set, we got a precision of 0.73. Thus, for any new customer, given an observation sequence, if the probability of churning is above 0.5, then we say the customer will churn, and otherwise they won’t.