Gibbs Sampling for Naive Bayes

Snehika Pandey[1,*]

[1]Data Science, Baltimore County, 21227, United States
[*]spandey3@umbc.edu

ABSTRACT

Gibbs Sampling technique implementation in News20group Dataset, sampling when it is difficult to sum up the integral

Introduction

Sampling is done to infer about the population, for sampling we must use inferential statistics which enables us to determine a population's characteristics by directly observing only a portion (or sample) of the population. Gibbs Sampling for Naive Bayes is what we focus in this paper

Sampling

Why Sampling?

Sampling can be used for the purpose of research and analysis on a population also making it easier to use and accurate representation of the population.

Sampling Methods

Sampling can be done in multiple ways: Some of the well known ways are Simple Random Sampling, Stratified Random Sampling, Systematic Sampling, Cluster Sampling and some more.

Two most basic sampling methods are:

- Inversion
- Rejection

Gibbs Sampling is one of Markov Chain Monte Carlo Method to approximate integrals, we need to approximate the integral in order to reduce the time consumed in working out. It is done by deriving posterior distribution and approximating it. Using Posterior Probability as a mode of likelihood has many benefits as compared to point estimates

Markov Chain Monte Carlo Sampling

Markov Chain Monte Carlo is a Bayesian Statistics Technique which is used for solving multivariate text analytics problem. Markov chain is a system which experiences transitioning from one state to another according to some probability rules, while Monte Carlo is a technique to approximate integrals, combining we can say that Markov chain Monte Carlo is a randomized algorithm which is use d to solve problems of complex population which are difficult to solve by integrals.

Gibbs Sampling

Gibbs Sampling is a Markov Chain Monte Carlo method which help us obtaining sequence when direct sampling is difficult to obtain. The idea is to find unknown variable conditioned on other variable, Latent Dirichlet Analysis is used to implement Gibbs Sampling. The sequence generated from Gibbs sampling can be used to approximate the joint distribution. Gibbs sampling draw prior from Dirichlet distribution and observed value from multinomial distribution.

Naive Bayes

Naive Bayes models have been widely used for clustering and classification. However, they are seldom used for general probabilistic learning and inference (i.e., for estimating and computing arbitrary joint, conditional and marginal distributions).

Latent Dirichlet Analysis

In LDA, each document may be viewed as a mixture of various topics where each document is considered to have a set of topics that are assigned to it via LDA. This is identical to probabilistic latent semantic analysis (pLSA), except that in LDA the topic distribution is assumed to have a sparse Dirichlet prior. The sparse Dirichlet priors encode the intuition that documents cover only a small set of topics and that topics use only a small set of words frequently. In practice, this results in a better disambiguation of words and a more precise assignment of documents to topics. LDA is a generalization of the pLSA model, which is equivalent to LDA under a uniform Dirichlet prior distribution.

With plate notation, which is often used to represent probabilistic graphical models (PGMs), the dependencies among the many variables can be captured concisely. The boxes are "plates" representing replicates, which are repeated entities. The outer plate represents documents, while the inner plate represents the repeated word positions in a given document, each of which position is associated with a choice of topic and word. M denotes the number of documents, N the number of words in a document. The variable names are defined as follows:

alpha is the parameter of the Dirichlet prior on the per-document topic distributions, beta is the parameter of the Dirichlet prior on the per-topic word distribution, theta is the topic distribution for document, and are also considered as hyper-parameters

Algorithm

The Naive Bayes (NB) model is fairly straightforward, because it assumes all of its features are independent of each other. In this example, we're labeling documents into one of two classes. To do this, we use Bayes' rule to flip the script, instead looking at a generative story

Given a bag of words(no sequence) of unlabeled document we generate labels using Bayes Rule, Implementing Gibbs sampler for naïve bayes demand efficient employment of conjugate priors and how we sample from conditional distribution. We try to generate document by considering words as features and sentiment label is either 0 or 1.

References

[1] Andrieu, Freitas, Doucet, and Jordan. An introduction to MCMC for machine learning. Machine Learning, 50:5–43, 2003.
[1] Philip Resnik, and Eric Hardisty. GIBBS SAMPLING FOR THE UNINITIATED, June 2010.