DEEP LEARNING

MOTIVATION FOR SAMPLING, GIBBS SAMPLING FOR TRAINING RBMs, CONTRASTIVE DIVERGENCE FOR TRAINING RBMs

Session - 32

CO - 4

AIM

To familiarize the students with the concepts of sampling and the types of sampling techniques for training RBMs

INSTRUCTIONAL OBJECTIVES

This session is designed:

To introduce the concepts of sampling and various types of sampling techniques for training RBMs

LEARNING OUTCOMES

At the end of this session, students will be able to:

Learn the concepts of sampling and the types of sampling techniques such as Gibbs sampling and Contrastive Divergence sampling for training RBMs

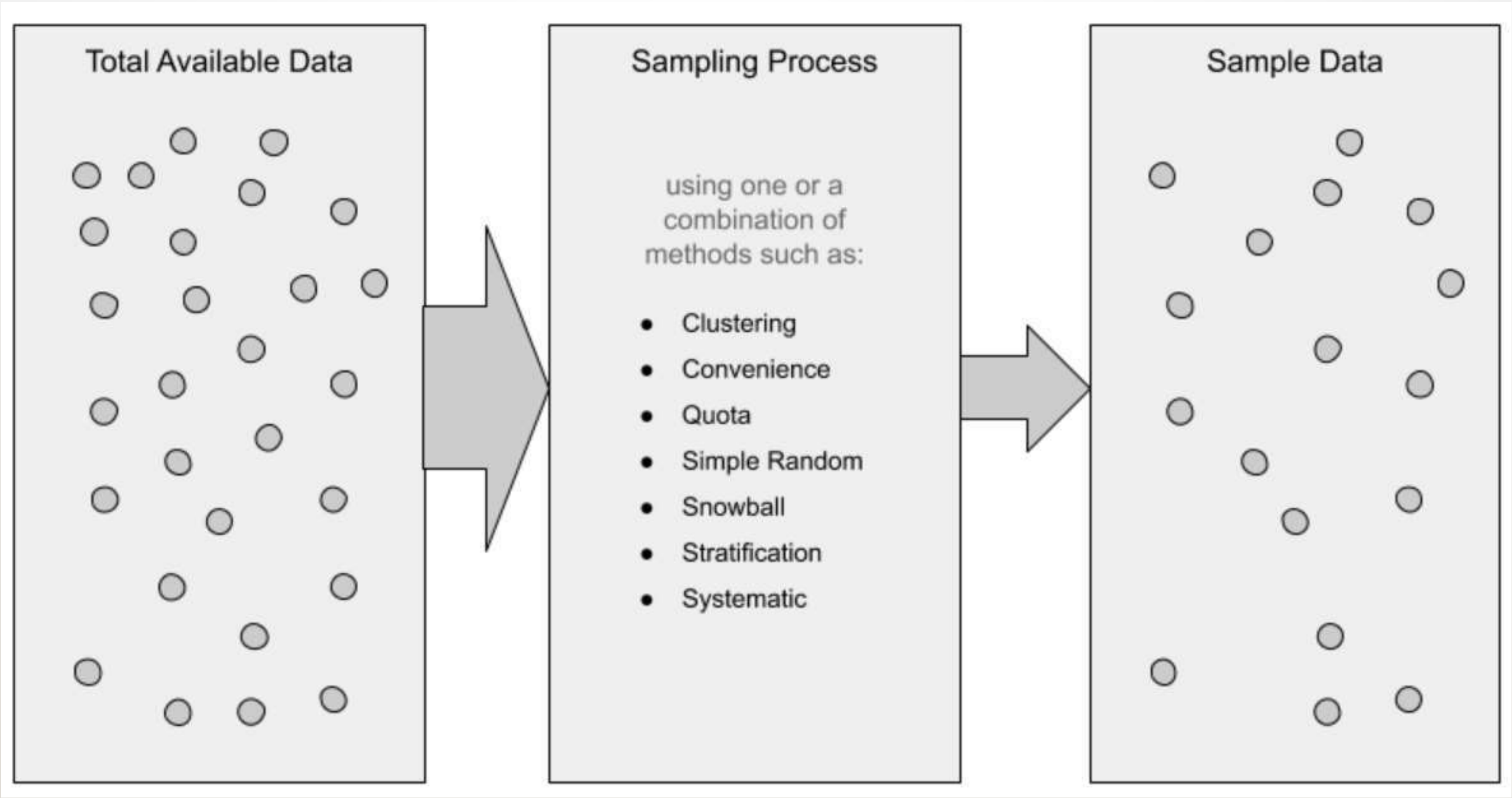
OUTLINE

- Motivation for sampling
- Gibbs sampling for training RBMs
- Contrastive divergence sampling for training RBMs
- Summary

MOTIVATION FOR SAMPLING: WHAT IS SAMPLING?

- Sampling in deep learning refers to the process of randomly selecting a subset of data points from a larger dataset.
- It is a common technique used for various purposes in machine learning and deep learning, including training, validation, and testing.

MOTIVATION FOR SAMPLING: SAMPLING PROCESS



MOTIVATION FOR SAMPLING

- Sampling plays a crucial role in deep learning for various reasons. The following are some key motivations for using sampling in deep learning.
  - Efficiency in Training
  - Data Augmentation
  - Imbalanced Datasets
  - Bootstrapping
  - Exploration and Uncertainty Estimation

MOTIVATION FOR SAMPLING

- Efficiency in Training
  - Sampling allows to work with smaller, manageable subsets of the data, making training faster and more practical
- Data Augmentation
  - In image classification and computer vision tasks, data augmentation involves applying random transformations (e.g., rotations, flips, cropping) to the training data.
  - These augmented samples help the model generalize better and become more robust to various inputs.
- Imbalanced Datasets
  - Sampling techniques (oversampling or undersampling) can be applied to balance the class distribution. This helps prevent the model from being biased towards the majority class

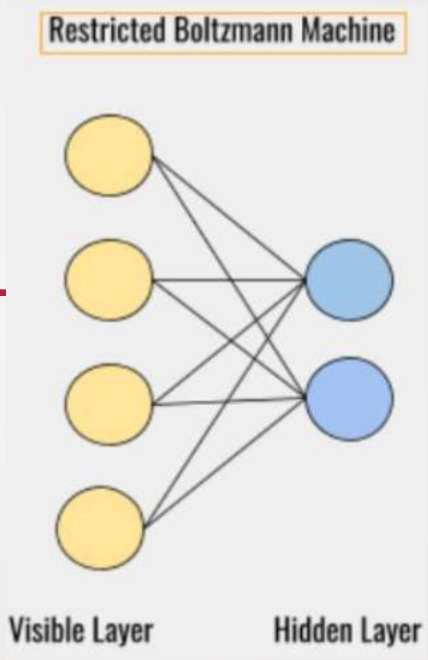
MOTIVATION FOR SAMPLING

- Bootstrapping
  - Bootstrap samples can be used to train multiple models (ensemble learning), leading to better performance and model robustness.
- Exploration and Uncertainty Estimation
  - In reinforcement learning and Bayesian deep learning, sampling is used to explore action spaces and estimate uncertainty.

TRAINING OF RESTRICTED BOLTZMANN MACHINE

- The training of the Restricted Boltzmann Machine differs from the training of regular neural networks via stochastic gradient descent.
- The two main training steps are:
  - Gibbs Sampling
  - Contrastive Divergence step

GIBBS SAMPLING FOR TRAINING RBMs



- **Initialization:**
  - Start with an initial configuration of the visible and hidden units.
  - These initial values can be randomly set or chosen based on the data.
- **Gibbs Sampling Steps:**
  - a. **Update Hidden Units:**
    - Given the current state of the visible units, sample new states for the hidden units from their conditional distribution.
    - Each hidden unit's state is independently sampled based on the probability that it is activated, which is a sigmoid function of the weighted sum of its connections to the visible units.
  - b. **Update Visible Units:**
    - Given the updated hidden unit states, sample new states for the visible units from their conditional distribution.
    - Again, each visible unit's state is independently sampled based on the probability that it is activated, using a sigmoid function of its connections to the hidden units.

GIBBS SAMPLING FOR TRAINING RBMs

- **Repeat the Steps, Update Hidden Units and Update Visible Units:**
  - Iterate through the Gibbs Sampling steps for a certain number of iterations (or until convergence is reached).
  - Each iteration constitutes one Gibbs Sampling step, and the states of both visible and hidden units are updated accordingly.
- **Collect Samples:**
  - After a sufficient number of Gibbs Sampling steps, the state of the RBM has evolved to a sample from the model's probability distribution.
  - This sample can be used for various tasks, such as generative modeling, feature learning, or inference.

CONTRASTIVE DIVERGENCE SAMPLING FOR TRAINING RBMs

- **Contrastive Divergence (CD)** is a popular training algorithm for Restricted Boltzmann Machines (RBMs) that uses a form of Gibbs Sampling to approximate the gradient of the log-likelihood of the data.
- The key idea is to perform a few steps of Gibbs Sampling to estimate the positive and negative phases needed for updating the RBM's parameters (weights and biases).
- CD-k, where k is the number of Gibbs Sampling steps, is a widely used variant of this algorithm.

CONTRASTIVE DIVERGENCE SAMPLING FOR TRAINING RBMs

- Initialization: Initialize the RBM's parameters, including the weights (connection strengths between visible and hidden units) and biases (one for each visible and hidden unit). Typically, these parameters are initialized randomly.
- Positive Phase:
  - Given a training sample (a configuration of visible units), compute the activations of the hidden units.
  - These activations are calculated as the sigmoid of the weighted sum of visible units connected to each hidden unit.
  - Sample the hidden units' states from their conditional distribution using the activations.
  - Compute the positive gradient term, which is the outer product of the visible and hidden unit activations.
- Negative Phase (CD-k):
  - Perform k steps of Gibbs Sampling. In each step:
    - Given the current state of the hidden units, sample new states for the visible units from their conditional distribution.
    - Given the updated visible unit states, sample new states for the hidden units from their conditional distribution.
  - After k Gibbs Sampling steps, the model has reached a new state (a "fantasy sample").
  - Compute the negative gradient term, which is the outer product of the visible and hidden unit activations from this fantasy sample.

CONTRASTIVE DIVERGENCE SAMPLING FOR TRAINING RBMs

- Update Parameters:
  - The RBM's parameters are updated by taking a step in the direction that minimizes the difference between the positive and negative gradient terms.
  - The specific learning rate and update rule depends on the optimization algorithm used (e.g., stochastic gradient descent).
- Repeat: **Steps, positive phase, negative phase, and update parameters are repeated** for a batch of training samples or the entire training dataset for several epochs until the RBM converges to a suitable model.
- Choice of k in CD-k:
  - The choice of the number of Gibbs Sampling steps, k, is an important consideration. In practice, small values of k (e.g., 1 or 2) are often sufficient for CD-k.
  - Increasing k can lead to more accurate gradient estimates but also increases computational cost.

SUMMARY

- Restricted Boltzmann Machines (RBMs) are probabilistic graphical models used in machine learning for unsupervised learning tasks.
- Sampling methods like Gibbs sampling, particularly in the form of contrastive divergence, are essential tools for training RBMs by approximating the likelihood gradients and updating model parameters to capture the underlying structure of the data.

BOOKS

Text Books:

- Ian Goodfellow and Yoshua Bengio and Aaron Courville, Deep Learning Book, MIT Press (2016)
- Grokking Deep Learning ,Andrew Trask, Manning publications, 2019
- Deep Learning with Python, Francois Chollet, Manning publications, 2018

Reference Books :

- Deep Learning with PyTorch: A practical approach to building neural network models using PyTorch by Vishnu Bramanian
- Neural Networks: A Systematic Introduction, Raúl Rojas, 1996 Pattern Recognition and Machine Learning, Christopher Bishop, 2007

WEB LINKS

https://www.linkedin.com/pulse/gibbs-sampling-training-rbms-jeevitesh-baluvu/

https://www.analyticssteps.com/blogs/what-restricted-boltzmann-machine-gibbs-sampling-and-contrastive-divergence

Self-Assessment Questions

1. Choose the correct option regarding the sampling.
   A. the sample is the population's part
   B. it helps in determining sampling error
   C. sampling saves money, time, and energy
   D. all these options are correct

2. What is the purpose of sampling in RBM training?
   A. Generate random data for training
   B. To approximate the probability distribution of visible and hidden units
   C. To calculate the gradient of the cost function
   D. To select random training examples

3. In RBM training, what is the objective of sampling in the Contrastive Divergence algorithm?
   A. To estimate the gradient of the log-likelihood
   B. To minimize the energy of the RBM
   C. To maximize the likelihood of the data
   D. To calculate the Hessian matrix

TERMINAL QUESTIONS

1. Explain the basic idea behind Gibbs sampling in the context of training RBMs.
2. How does Gibbs sampling help in approximating the likelihood gradient for RBM training?
3. What role does the contrastive divergence algorithm play in the context of Gibbs sampling and RBMs?
4. Discuss the impact of choosing different numbers of Gibbs sampling steps on the convergence and efficiency of RBM training.

THANK YOU

Team – Deep Learning