**Introduction**

In this assignment, you will train a **Restricted Boltzmann Machine** to perform **Topic Modelling** on a set of Amazon reviews.

Let's start by understanding what Topic Modelling is.

Topic Modelling is the art and science of identifying the 'latent topics' in a text. It is an unsupervised problem. You input a set of documents/ corpus into the model and the model finds the topics that describe the corpus. Each topic is a distribution over the words that best describe the topic.

 Let's understand this with the help of an example:

Suppose you are given reviews about a product, for example, iPhone. You can perform topic modelling to see what topics are the people in the reviews talking about. Consider that we have input to the model that we want 3 top topics from the reviews.

 The top 5 words that describe the topics are shown below:

|  |  |  |
| --- | --- | --- |
| **Topic 1** | **Topic 2** | **Topic 3** |
| pixel | smooth | slow |
| selfie | apps | memory |
| good | hangs | storage |
| blur | interface | full |
| filter | intelligent | apps |

Looking at the above words, we can say that Topic 1 mostly talks about the camera of iPhone while Topic 2 talks about the iOS operating system and Topic 3 talks about the storage memory available. It might often happen that you might not be able to figure out the topics from the distribution of the words.

 After getting the topics, you can do a lot of things like:

1. Find the distribution of each review over the topics
2. Couple the topics and sentiments of each review and see which topics more often in negative reviews
3. Use the topic representation of each document as a feature representation for developing further machine learning models

You can use different models to perform topic modelling, one of which is LDA  which you have studied earlier in the Natural Language Processing course.

In this assignment, we will use a Restricted Boltzmann Machine (RBM) to perform topic modelling. The input to an RBM is a bag of words model and the output is the distribution of words in the topics.

The motivation behind studying RBMs is that these are building blocks of **Deep Belief Networks**, a generative class of models which have become quite popular lately.

# Structure of RBM

An RBM has a neural network equivalent structure with a visible layer and a hidden layer. Just like a neural network, there are weights associated with the connection between the layers. You also learnt that an RBM is a bipartite structure. A bipartite structure is one where none of the nodes in a particular layer is connected to each other.

Let's understand the structure of an RBM as a probabilistic graphical model.

Using the concepts of independence learned earlier, the nodes in the hidden layer are independent of each other conditioned on all the nodes of the visible layer.  Similarly, the nodes in the visible layer are independent of each other conditioned on all the nodes of the hidden layer.

Let's first define the joint probability distribution using the energy function in the next segment.

# Energy Function in RBM

 Joint probability distribution p(v,h) as follows:



where E(v,h) is the energy function and Z is the normalizing factor.

Let's understand the structure and the functioning of an RBM as a neural network.

Using a neural network scheme of calculation with all the neurons having the sigmoid activation function

p(hj=1|V)=σ(cj+∑ni=1wijvi)

p(vi=1|H)=σ(bi+∑mj=1wijhj)

The above results of a neural network which have been represented as a conditional probability are consistent with the PGM view of RBMs when we use the following energy function:

E(v,h)=∑wijvihj+∑bivi+∑cjhj

 Let's now try to delve into the functioning of an RBM in a little more depth.

# Functioning of RBMs

Before we make use of this energy function to train the RBM, let's understand the functioning of an RBM in a little more depth.

You have seen that an RBM functions like an autoencoder where the hidden layer creates the representation of the input layer upon training. An autoencoder creates a meaningful feature representation of the input in such a way that you will be able to recreate the image using its feature representation.

 Hence, the training method here is that, given a set of inputs which are binary variables, what are the values of weights and biases such that the joint probability distribution is maximised or the energy is minimized.

You have already seen what a bag of words model is and how to create one in the Natural Language Processing course. In the bag of words model, each row of the matrix becomes the visible vector representation of the document. Hence, by passing the bag of words model at once, you'll be able to train the RBMs in batches of data points.