CS 247 : Advanced Data Mining Learning **(Due: 11:59 pm 04/24/19)**

Homework Assignment #2

Instructor: Yizhou Sun *TA:* Zeyu Li

Homework Policy:

- Read the ***Homework Submission Guidance*** carefully before you start working on the assignment, and before you make a submission.
- This is an individual homework. Please **DO NOT** collaborate with others.
- Write your answers in the notebook `cs247_hw2.ipynb`. Only submit a **pdf** converted from `cs247_hw2.ipynb` to Gradescope.

Problem 1: Multinomial Mixture Model

Similar to Gaussian Mixture Model, Multinomial Mixture Model can also be used for soft clustering. For example, in the case of document clustering, each cluster k is a categorical distribution with parameters β_k, and $x_i|z_i = k \sim Categorical(\beta_k)$, where x_i is a bag-of-words representation vector for document i. Please write down the EM algorithm for soft document clustering under Multinomial Mixture Model.

Problem 2: Semi-supervised learning

Suppose first l documents in a corpus are observed with labels and the remaining ones are without labels, i.e., $D = \{x_i, y_i\}_{i=1}^{l} \cup \{x_i\}_{i=l+1}^{n}$. How can we use both labeled documents and unlabeled documents together to learn a document classifier? One way is to write down the likelihood for the dataset as a combination for two parts, where the first part models $p(x_i, y_i)$ for labeled documents and second part models $p(x_i)$ for unlabeled documents.

1. Please write down the model and the learning algorithm. (Hint: Fix the labels for the first l documents. In the *E-step*, only unlabeled documents are involved; In the *M-step*, both labeled and unlabeled documents will be used for parameter update.)
2. Please implement the algorithm, and compare the classification results with purely supervised method (i.e., naïve Bayes in HW1), where the unsupervised component from the model is dropped, for test documents. For this task, we will play with a sentiment analysis dataset included in the handouts (`p2.labeled` and `p2.unlabeled`). Detailed descriptions are in the notebook file and here.

Problem 3: Extending logistic regression by introducing hidden layers.

1. For a dataset with three data points $x^{(1)} = (-1, 2)$, $x^{(2)} = (0, 2)$, and $x^{(3)} = (1, 2)$, and labels $y^{(1)} = 1$, $y^{(2)} = 0$, and $y^{(3)} = 1$, can logistic regression separate them?
2. If we add another hidden layer into the logistic regression corresponding to the neural network architecture, can some of these neural networks separate these data points into two classes? Please give an example of such neural network that can perform the task (weights need to be specified).
3. Compare the performance of logistic regression and the neural network with one hidden layer that extends logistic regression on the Australia Rain dataset. This dataset contains daily weather data from a number of Weather Stations. Here are some detailed information about the dataset:
   - The target to predict is `RainTomorrow`.
   - Remove the `RiskMM` feature as it is a leak to the answer to be predicted.
   - Feel free to do normalization, standardization, or other preprocessing on existing features.
   - More information or statistics about the dataset is here.