**Machine Learning & Data Mining, Spring 2020**

**Homework 11**

Due June 14

* True or False (Explain your answer).

1. To train an RNN model, the length of input sequences in the training data must be the same.

=> F

RNN 모델은 Sequential data에 적합하기 위해 개발된 모델로서, 가변적인 input길이에 대응할 수 있도록 설계되어, 이전 단어가 다음 단어에 영향을 끼치고(이전 hidden layer state가 다음 hidden layer state에 영향 끼치는 것으로 가상화) many to many, many to one, one to many, many to many의 sequence modeling이 가능하기에 입력 길이는 가변적이어도 문제가 없습니다.

1. GRU has more parameters to learn than LSTM does.

=> F

LSTM과 GRU는 GATE를 사용하여 이전 정보 흐름량을 조절하는 RNN의 확장모델인데 차이는 사용 GATE의 종류와 수 입니다. LSTM은 FORGET, INPUT, OUTPUT 3가지 GATE를 사용, GRU는 UPDATE(FORGET + INPUT), RESET GATE 2가지를 사용합니다. GATE도 하나의 Neural Network이고, LSTM과 비교시 GRU는 하나의 GATE가 더 적으므로, GRU가 더 많은 파라미터를 학습한다는 것은 틀린명제입니다.

1. K-means clustering works better for a large sized dataset than a hierarchical clustering algorithm.

=> T

Hierarchical clustering은 클러스터의 갯수를 정하지 않고 시작하기에 유연하다는 장점이 있으나 linkage 알고리즘이 pointwise하게 모든 datapoint를 비교하기 때문에, 결과적으로 O(이라는 시간복잡도를 가지게 되고, 반면에 K means clustering은 반복횟수 l, feature 수 m, centroid의 갯수 k, sample 수 n이라고 할 때, O(knml)이라는 시간복잡도를 갖는데, 이는 모든 요소에 linear하게 변하는 복잡도로서 데이터크기 n이 지대한 영향을 끼치는 데이터에서는 k means가 더 좋은 선택입니다.

1. The choice for initial centroids does not affect the K-means clustering result.

=> F

알고리즘의 과정을 간략화하면 centroid를 기준으로 점들이 모이고, centroid가 업데이트 되는데, 처음에 모이는 것에 따라 완전히 다른 모양이 나옵니다. 때문에 초기 centroid가 영향을 안 끼친다는 것은 잘못된 명제입니다.

1. The choice for the linkage method in hierarchical clustering does not affect the clustering result.

=> F

max, average, min 등의 linkage 알고리즘이 있는데 이 것들은 클러스터 사이의 거리를 재는 척도로 사용되고, 서로 다른 기준을 사용하여 거리를 잽니다. 기준이 달라 거리개념이 완전히 달라지기 때문에 당연히 결과에 영향을 끼칩니다. 때문에 linkage method가 클러스터링 결과에 영향을 안끼친 다는 것은 틀린 명제입니다.

1. Spectral clustering is a graph clustering algorithm and cannot be applied to a typical data matrix (sample-feature matrix).

=> F

input이 그래프 형태로 들어와야 하는데, sample-feature vector또한 distance matrix 등을 이용하여 그래프화 할 수 있기에 spectral cluster의 입력으로 사용될 수 있습니다.

1. If we assume a linear activation function in Autoencoder, the result will be the same as the one from PCA.

=> f

AU의 ACTIVATION함수가 LINEAR이고, 손실함수로 MSE를 쓰고 node수를 줄이는 압축의 개념 등 제약 조건을 만족한다면 같다.

1. The loss function in t-SNE is based on the difference between two probability distributions.

=> T

t-SNE는 t 분포를 기반으로 두 확률분포의 거리 차이를 정의할 수 있는 정보이론기법인 KL-divergence를 사용합니다.