HỌC MÁY

BÀI TẬP GIỮA KỲ

**Bài 1:**

1. Trình bày về mục tiêu, ý nghĩa của nhiệm vụ Word2Vec (còn được hiểu là sinh Word Embedding)
2. Trình bày cách huấn luyện để tạo vector từ words đối với phương pháp BOW và Skip-Gram
3. Hãy trình bày các phương pháp Word2Vec khác: Fasttext, Glove
4. Sử dụng thư viện Gensim để huấn luyện mô hình word2vec cho tiếng Việt
5. Hãy tìm hiểu các cách biểu diễn Vector của 1 câu (Sentence2Vec), sử dụng để phân loại câu hoặc đo độ tương tự của 2 câu

**Bài 2:**

1. Tìm hiểu bài toán phân loại văn bản đa nhãn và ứng dụng thuật toán dựa trên Neural Network để giải quyết bài toán này.
   1. Sử dụng Feed Forward Neural Network
   2. Sử dụng các loại RNN (Recurrent Neural Networks)

Lưu ý: có thể thử nghiệm cả 2 phương án, chọn hoặc ko chọn việc sử dụng Word2vec hoặc Sentence2Vec

1. Vẫn bài toán ở câu 1 nhưng sử dụng các phương pháp truyền thống: SVM, Logistic Regression, Naive Bayes classification.

So sánh tất cả các phương pháp, bao gồm cả các phương pháp ở câu 1