# Báo cáo môn Trí tuệ nhân tạo: Phân tích cảm xúc (vui vẻ, buồn) trên 1 tập tweet bằng logistic regression

## 1. Thu thập và xử lý dữ liệu

* Dữ liệu lấy từ tập dữ liệu ntlk.corpus.twitter\_samples, gồm 1 list các tweet, ví dụ: RT @Twitter @chapagain Hello There! Have a great day. #good #morning <http://chapagain.com.np>
* Loại bỏ RT, URL, hastag trong các tweet

## 2. Trích xuất các feature

* Các tweet được biểu diễn bởi bias 2 feature: số lượng các từ positive và số lượng các từ negative

## 3. Logistic Regression

### 3.1 Hàm Sigmoid

* Hàm sigmoid chuyển giá trị trong khoảng thành

### 3.2 Gradient Descent

* Trọng số mô hình:
* Input x có chiều (m, 3) với m là số sample
* Dự đoán:
* Hàm lost:

* Cập nhật theta: