Các khái niệm:

* Loss function
* Gradient Descent / Stochastic gradient descent
* Learning rate
* Mini batch, epoch
* Active function
* Soft-max function
* Training, test, validation, data sets
* Cross validation test
* Precision, recall, f score, accuracy
* Overfitting and regularization
* Dropout

Loss function:

* Thể hiện sự chênh lệch giữa 2 đại lượng: label được dự đoán và label đúng
* Hàm mất mát có giá trị nhỏ khi phép đánh giá cho kết quả tốt và ngược lại.
* Mục tiêu là tìm các tham số mô hình để hàm mất mát có kết quả nhỏ nhất.

Gradient Descent

* Trong machine learning, ta thường xuyên phải tìm giá trị lớn nhất và nhỏ nhất của một hàm số.
* Việc giải phương trình đạo hàm bằng không để tìm cực tiểu là bất khả thi.
* Hướng tiếp cận: Xuất phát từ một điểm được coi là gần nhất với nghiệm, sau đó lặp lại để tiến dần đến điểm cần tìm (đạo hàm gần bằng không).
* Giả sử tìm được điểm x sau vòng lặp t, ta cần tìm thuật toán đưa x về càng gần cực tiểu càng tốt.
  + Nếu đạo hàm tại x dương ( > 0 ), thì x nằm bên phải cực tiểu hoặc ngược lại. Cần di chuyển x về bên trái, phía âm, tức là ngược dấu với đạo hàm
  + n: tốc độ học (learning rate)

Stochastic Gradient Descent