Họ và tên: Đặng Anh Tiến
MSSV: 20520800
Lớp: CS431.N11

Bài tập LT2: Giải thích hàm loss function của Logistic Regression

Trong bài toán Binary Classification, miền giá trị của y_pred(kết quả dự đoán) nằm trong khoảng (0, 1); y_true(kết quả thực tế) là giá trị 0 và 1.

- Trong trường hợp kết quả dự đoán gần với kết quả thực tế (y_true == y_pred): y_true = 0, y_pred~0

  MSE = (0 - 0)^2 = 0
  BCE = -(0 * log(0) + log(1)) = 0

  => Cho thấy khi mô hình dự đoán đúng, kết quả của hai hàm loss không có quá nhiều sự khác biệt.

- Trong trường hợp kết quả dự đoán khác với kết quả thực tế (y_true != y_pred): y_true = 0, y_pred ~ 1

  MSE = (0 - 1)^2 ~ 1
  BCE = -(0 * log(1) + log(0)) = -(-inf) = +inf

  => BCE trả về +inf cao hơn rất nhiều so với MSE.

Có thể thấy trong bài toán Binary Classification, ta nên dùng Binary Cross Entropy (BCE) thay vì Mean Squared Error (MSE) bởi vì các lý do sau:

1. Giá trị của y_true và y_pred trong khoảng [0, 1], vì vậy MSE luôn trả về giá trị trong khoảng [0, 1]. Khi model dự đoán sai,BCE, như đã chứng minh ở trên, trả về kết quả lớn hơn rất nhiều (+inf) so với MSE. Điều này khiến BCE phạt hàm Loss nặng hơn.
2. Giá trị hàm Loss lớn hơn, gradient lớn hơn, vì vậy trong Gradient Descent, tối ưu các parameter sẽ diễn ra nhanh hơn. MSE có giá trị trong [0, 1], các gradient sẽ có giá trị [0, 1], Gradient Descent sẽ tối ưu vô cùng chậm. Trong khi đó, BCE cho kết quả Loss lớn, gradient sẽ lớn, điều này giúp Gradient Descent tối ưu nhanh hơn nhiều so với MSE.