

BT9 - Hàm loss cho classification head và regression head

Hoàng Tiến Dũng - MSSV: 19521388

1 Bài toán regression và classification khác nhau như thế nào?

Cơ sở	Classification	Regression
Bản chất	Classification là bài toán dự đoán các giá trị rời rạc	Regression là bài toán dự đoán các giá trị liên tục
Dự đoán giá trị	Giá trị rời rạc, không có thứ tự	Giá trị liên tục, có thứ tự
Thuật toán	Logistic regression, KNN, Soft-max regression,...	Linear regression, Random forest,...
Đánh giá	Accuracy	Root mean squared error

2 Hàm lỗi cho bài toán classification và regression thường là hàm gì? Tại sao?

- Hàm lỗi của Regression thường là Mean Square Error(MSE).
MSE như tên gọi của nó là tính trung bình của bình phương chênh lệch giữa các giá trị dự đoán và giá trị thực sự. MSE phù hợp với bài toán Regression là vì output của bài toán regression là các giá trị liên tục và có thứ tự tức là sau mỗi lần cập nhật trọng số khi sử dụng gradient descent thì output tiến dần về giá trị thực hơn.
- Hàm lỗi của bài toán Classification thường là Cross Entropy Loss.
Cross Entropy được xem là một hàm loss tốt cho Classification, bởi vì nó giảm thiểu khoảng cách giữa hai phân phối - dự đoán và thực tế. Bài toán classification dự đoán các giá trị rời rạc và không có thứ tự. Output của bài toán chính là xác suất của mỗi giá trị rời rạc ấy. Ví dụ trong bài toán phân loại mèo, chó, gà: một bức ảnh của con mèo được model dự đoán là $[0.2, 0.7, 0.1]$ trong khi xác suất thực sự là $[1, 0, 0]$. Ý tưởng của cross entropy là giảm thiểu sự khác biệt giữa hai xác suất này.