

Basic Python - Einsum in Numpy

Hoàng-Nguyên Vũ

Confusion Matrix và Các Metric Đánh Giá

Confusion matrix và các metric đánh giá liên quan là những công cụ quan trọng trong học máy (machine learning) để đánh giá hiệu suất của mô hình phân loại. Dưới đây là mô tả chi tiết về confusion matrix và các metric đánh giá phổ biến:

1. Confusion Matrix

Confusion matrix là một bảng ma trận có kích thước $m \times m$ (với m là số lượng lớp) được sử dụng để mô tả hiệu suất của một mô hình phân loại. Mỗi hàng của ma trận đại diện cho các mẫu thuộc lớp thật (actual class), trong khi mỗi cột đại diện cho các mẫu được mô hình dự đoán (predicted class).

		True Class	
		Positive	Negative
Predicted Class	Positive	TP	FP
	Negative	FN	TN

Một confusion matrix cơ bản cho bài toán phân loại nhị phân (2 lớp) bao gồm 4 thành phần chính:

- **True Positives (TP)**: Số lượng mẫu mà mô hình dự đoán đúng là dương.
- **True Negatives (TN)**: Số lượng mẫu mà mô hình dự đoán đúng là âm.
- **False Positives (FP)**: Số lượng mẫu mà mô hình dự đoán là dương nhưng thực tế là âm (còn gọi là Type I error).
- **False Negatives (FN)**: Số lượng mẫu mà mô hình dự đoán là âm nhưng thực tế là dương (còn gọi là Type II error).

2. Các Metric Đánh Giá

Từ confusion matrix, chúng ta có thể tính toán các metric đánh giá hiệu suất của mô hình, bao gồm:

Accuracy (Độ chính xác):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Đây là tỉ lệ phần trăm các mẫu được dự đoán đúng trong tổng số mẫu.

Precision (Độ chính xác dương):

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

Đây là tỉ lệ mẫu dương thật trong số các mẫu được dự đoán là dương.

Recall (Độ nhạy):

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

Đây là tỉ lệ mẫu dương thật được nhận diện đúng bởi mô hình.

F1 Score:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

F1 Score là trung bình điều hòa giữa Precision và Recall, giúp cân bằng giữa hai metric này.

Specificity (Độ đặc hiệu):

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (5)$$

Đây là tỉ lệ mẫu âm thật được nhận diện đúng bởi mô hình.

3. Áp Dụng và Ý Nghĩa

- **Accuracy** là một chỉ số tổng quát, nhưng có thể gây hiểu nhầm nếu dữ liệu không cân bằng (chẳng hạn, khi số lượng mẫu thuộc một lớp nhiều hơn lớp còn lại).
- **Precision và Recall** thường được sử dụng trong các trường hợp cần quan tâm đặc biệt đến các mẫu dương, như trong phát hiện bệnh tật (dương tính là có bệnh).
- **F1 Score** hữu ích khi muốn cân bằng giữa Precision và Recall, đặc biệt trong trường hợp dữ liệu không cân bằng.

- **Specificity** quan trọng khi muốn đảm bảo rằng các mẫu âm không bị nhầm lẫn với các mẫu dương.

Các metric này giúp cung cấp cái nhìn toàn diện về khả năng của mô hình phân loại, từ đó có thể điều chỉnh và cải thiện mô hình cho phù hợp với bài toán thực tế.

4. Bài tập

Cho một tập ảnh gồm 1000 ảnh, trong đó có 400 ảnh là ảnh mèo và 600 ảnh còn lại là ảnh không phải mèo (là ảnh con chó, ảnh tivi, ảnh cây cối...) Xây dựng một cỗ máy tìm kiếm (search engine) để tìm kiếm ảnh theo từ khóa nhập vào. Khi mình gõ chữ “mèo” để tìm kiếm ảnh “mèo” thì kết quả cỗ máy tìm kiếm trả về 100 ảnh, trong đó có 98 ảnh mèo và 2 ảnh không phải mèo.

Câu hỏi 1: Hãy xác định TP, TN, FP, FN dựa theo thông tin dưới đây:

- 98 ảnh mèo (trong 100 ảnh do search engine trả về) được gọi là gì?
- 2 ảnh không phải mèo (trong 100 ảnh do search engine trả về) được gọi là gì?
- 100 ảnh do search engine trả về được gọi là gì?
- 302 ảnh mèo còn lại (400 ảnh mèo – 98 ảnh mèo) được gọi là gì?
- 598 ảnh không phải mèo còn lại (600 ảnh không phải mèo – 2 ảnh không phải mèo) được gọi là gì?

Câu hỏi 2: Dựa trên câu hỏi 1, các bạn hãy tính precision, recall, f1 và đưa ra nhận xét.