**Báo cáo tuần 7: báo cáo bù về cách đánh giá độ chính xác của 1 model.**

1. **Overview**

* Khi xây dựng một mô hình Machine Learning, chúng ta cần một phép đánh giá để xem mô hình sử dụng có hiệu quả không và để so sánh khả năng của các mô hình.
* Hiệu năng của một mô hình thường được đánh giá dựa trên tập dữ liệu kiểm thử (test data). Điều này được thực hiện bằng cách cho tập kiểm thử là đầu vào 🡪 đầu ra (y\_predictions). Ta cần so sánh y\_predictions với các class thật của dữ liệu (y\_true)
* Có rất nhiều cách đánh giá một mô hình phân lớp. Tuỳ vào những bài toán khác nhau mà chúng ta sử dụng các phương pháp khác nhau.

1. **Accuracy (độ chính xác)**

* Đơn giản nhất: tính tỉ lệ giữa số điểm được dự đoán đúng và tổng số điểm trong tập dữ liệu kiểm thử.
* Ta có thể tự xây dựng hàm tính so sánh kết quả giữa dự đoán và thực tế hay có thể sử dụng hàm có sẵn ở các thư viện.
* VD: với bài toán phân loại 3 lớp dữ liệu có gán nhãn.

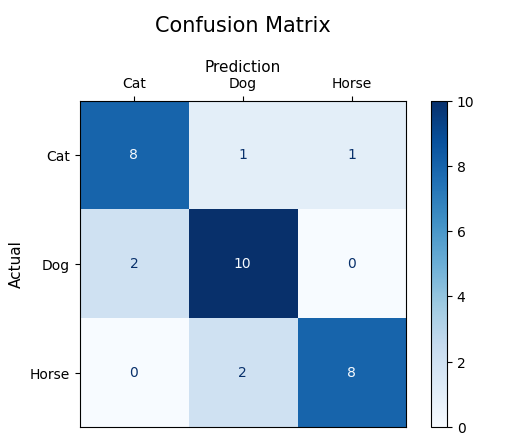
**A screenshot of a computer program

Description automatically generated**

* Nhược điểm: sử dụng acccuracy chỉ cho ta biết bao nhiêu phần trăm lượng dữ liệu được phân loại đúng mà không chỉ ra một số khía cạnh quan trọng khác ( độ chính xác của mỗi loại , dữ liệu thường bị phân nhần vào nhóm nào,...)

1. **Confusion matrix**

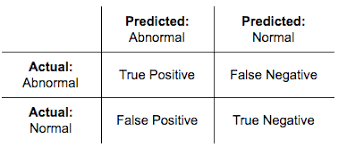
* Về cơ bản, confusion matrix thể hiện có bao nhiêu điểm dữ liệu thực sự thuộc về 1 class và được dự đoán rơi vào 1 class.
* Giá trị ô ở hàng thứ i cột j là số lượng điểm thực ra thuộc vào class i nhưng được dự đoán vào class j.
* Từ ma trận, ta có thể suy ra tổng số các phần tử là số điểm dữ liệu kiểm thử, các phần tử trên đường chéo chính là số điểm được phân loại/ dự đoán chính xác.



* Cách biểu diễn trên là confusion\_matrix chưa được chuẩn hóa. Để chuẩn hóa, ta có thể dử dụng normalized\_confusion matrix bằng cách lấy mỗi ô của ma trận trong một hàng chia cho tổng các phần tử trên hàng đó.
* Tổng các phần tử trên 1 hàng luôn bằng 1
* Confusion matrix thường được minh hoạ bằng màu sắc thông qua thư viện matplotlib để có cái nhìn rõ ràng hơn. Một mô hình tốt sẽ cho một confusion matrix có các phần tử trên đường chéo chính có giá trị lớn, các phần tử còn lại có giá trị nhỏ. Nói cách khác, khi biểu diễn bằng màu sắc, đường chéo có màu càng đậm so với phần còn lại sẽ càng tốt.

1. True/False Positive/Negative

* Đây là cách đánh giá quan trọng trong bài toán phân lớp có 2 lớp dữ liêu, trong 2 lớp có 1 lớp nghiêm trọng hơn lớp kia và cần được dự đoán chính xác.



* Ví dụ, trong bài toán xác định có bệnh ung thư hay không thì việc không bị sót quan trọng hơn là việc chẩn đoán nhầm âm tính thành dương tính. Các định nghĩa có thể hiểu trong bài toán này như sau:
* True Positive: dự đoán bệnh nhân bị ung thư và thực tế bệnh nhân bị ung thư. Đây là mảng ta cần có độ chính xác cao nhất để đảm bảo không bỏ sót người bị bệnh
* False Positive: dự đoán bệnh nhân bị ung thư nhưng thực tế ko bị
* True Negative: dự đoán bệnh nhân không bị ung thư và thực tế không bị thật
* False negative: dự đoán bệnh nhân không bị ung thư nhưng thật ra họ bị ung thư 🡪 trái với kết quả dự đoán, bỏ sót trường hợp bị bệnh, cần tối thiểu nhất có thể.
* Tương ứng với các thông số trên là các Rate dựa trên normalized confusion matrix. Ta đặc biệt quan tâm đến False Positive Rate (tỉ lệ báo động nhầm ) và False Negative Rate (tỉ lệ bỏ sót) . Trong bài toán phát hiện ung thư trên, thà báo nhầm còn hơn bỏ sót, tức là ta có thể chấp nhận False Alarm Rate cao để đạt được Miss Detection Rate thấp.

1. Precision và Recall

* Với bài toán phân loại mà tập dữ liệu của các lớp là chênh lệch nhau rất nhiều, có một phép đó hiệu quả thường được sử dụng là Precision-Recall.
* Precision được định nghĩa là tỉ lệ số điểm true positive trong số những điểm được phân loại là positive (TP + FP). Precision cao chứng tỏ mô hình đưa ra dự đoán đúng càng cao.

A black text on a white background

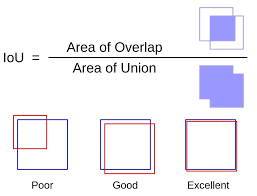
Description automatically generated

* Recall được định nghĩa là tỉ lệ số điểm true positive trong số những điểm thực sự là positive (TP + FN). Recall càng cao cho thấy mô hình ít có khả năng bỏ sót các đặc tính positive.

A black and white math equation

Description automatically generated with medium confidence

* Đi cùng cả 2 khái niệm precision- recall là precision-recall curve. Đường cong thể hiện relationship giữa precision/ recall khi thay đổi ngưỡng dự đoán (threshold) nào đó. Việc thay đổi ngưỡng nói trên ảnh hưởng đến cả precision và recall.
* Ngưỡng thấp 🡪 Nhiều data được dự đoán positive 🡪 recall cao nhưng precision thấp.
* Ngưỡng cao 🡪 Ít data được dự đoán positive 🡪 recall thấp nhưng precision cao.
* Trong bài toán object detection, precision và recall thường đi kèm với 1 thông số quan trọng khác là IOU: Intersection Over Unit. Khi một mô hình dự đoán một bounding box (hộp chứa đối tượng), chúng ta cần so sánh nó với ground truth box (hộp chứa đối tượng thực tế). IOU đo lường mức độ chồng lấp giữa hai hộp này.



* Thông thường, nếu IOU >= 0.5, predicted box sẽ được coi là TP.

1. AP, AR and mAP

* Average Precision (AP) là giá trị trung bình của precision khi thay đổi các ngưỡng khác nhau. AP đóng vai trò đo lường độ chính xác của mô hình trong việc phát hiện đối tượng.
* Average Recall (AR) là giá trị trung bình của recall khi thay đổi các ngưỡng khác nhau, đóng vai trò đo lường khả năng mô hình phát hiện tất cả các đối tượng có trong ảnh.
* Mean Average Precision (mAP) là giá trị trung bình của AP của tất cả các lớp đối tượng. Nó được tính bằng cách lấy trung bình cộng của AP của từng lớp đối tượng. mAP là một chỉ số tổng quát để đánh giá hiệu suất của mô hình phát hiện đối tượng.