

1. Приведите по 2 примера, когда лучше максимизировать Precision, а когда Recall.

Я думаю, что лучше максимизировать Precision при определении заболевания плода у беременной женщины (имеется ввиду серьезное заболевание, при котором необходимо прерывание беременности). В этом случае нам важна точность, чтобы не лишиться жизни здорового неродившегося ребенка.

Или, например, мы пытаемся как можно реже делать ложноположительные предсказания потому что это слишком дорого, как если мы ищем где копать нефть.

А Recall лучше максимизировать, когда определяем болен человек Covid-19 (или другим опасным и заразным заболеванием) или нет. В этом случае нам важна полнота, чтобы изолировать больного и недопустимо распространение инфекции.

Или, например, нужно определить порок сердца у человека. Тут лучше лишний раз допустить False Positive, и предположить болезнь у здорового человека, нежели упустить больного и не заметить ее.

2. Почему мы используем F-меру, почему, например, нельзя просто взять среднее от Precision и Recall?

F-мера представляет собой гармоническое среднее между precision и recall. Она стремится к нулю, если любой из этих параметров стремится к нулю. Эту метрику мы хотим видеть высокой если нам важен баланс.

Рассмотрим, например, всегда возвращающий класс A. Есть бесконечные элементы данных класса B и один элемент класса A:

Precision: 0.0 Recall: 1.0

Принимая среднее арифметическое, оно будет на 50% правильным. Несмотря на худший из возможных результатов! С гармоническим средним, мера F1 равна 0.

Arithmetic mean: 0.5 Harmonic mean: 0.0

Другими словами, чтобы иметь высокий F1, нужно иметь как высокую точность, так и полноту.

Например, что в среднем 30 миль в час и 40 миль в час?

Если вы едете в течение 1 часа на каждой скорости, средняя скорость за 2 часа - это действительно среднее арифметическое, 35 миль в час.

Однако, если вы проезжаете одно и то же расстояние на каждой скорости, скажем, 10 миль, то средняя скорость более 20 миль - это среднее гармоническое 30 и 40, около 34,3 миль в час.

Причина в том, что для того, чтобы среднее значение было действительным, вам действительно нужно, чтобы значения были в одинаковых масштабированных единицах. Мили в час нужно сравнивать за одно и то же количество часов; для сравнения того же количества миль вместо этого вам нужно усреднить часы на милю, что в точности соответствует значению гармонии.

Точность и полнота имеют положительные стороны в числителе и разные знаменатели. Чтобы усреднить их, на самом деле имеет смысл только усреднить их взаимные значения, то есть среднее гармоническое.