**METRICS**

Accuracy

Classification Accuracy is what we usually mean, when we use the term accuracy. It is the ratio of number of correct predictions to the total number of input samples.

https://miro.medium.com/max/373/1*yRa2inzTnyASJOre93ep3g.gif

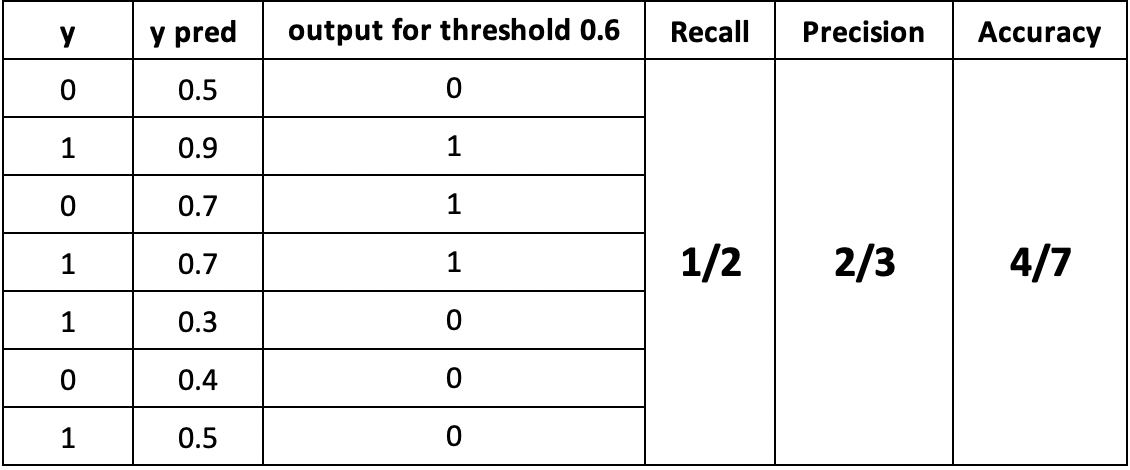
It works well only if there are equal number of samples belonging to each class.

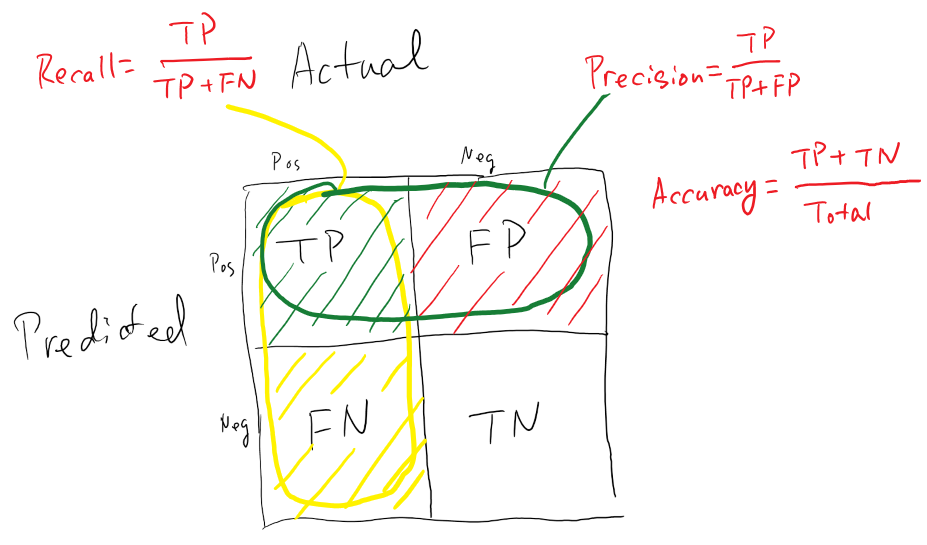
For example, consider that there are 98% samples of class A and 2% samples of class B in our training set. Then our model can easily get **98% training accuracy** by simply predicting every training sample belonging to class A.

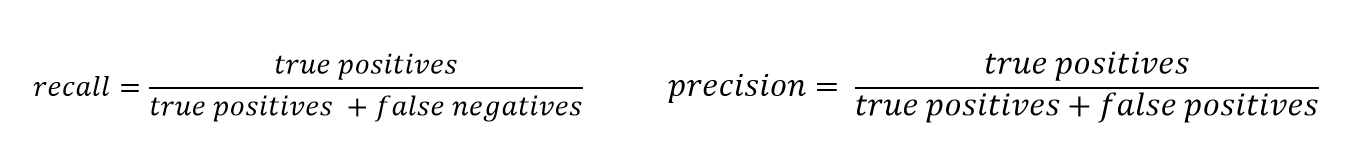
Confusion Matrix



* **True Positives** : The cases in which we predicted YES and the actual output was also YES.
* **True Negatives** : The cases in which we predicted NO and the actual output was NO.
* **False Positives** : The cases in which we predicted YES and the actual output was NO.
* **False Negatives** : The cases in which we predicted NO and the actual output was YES.

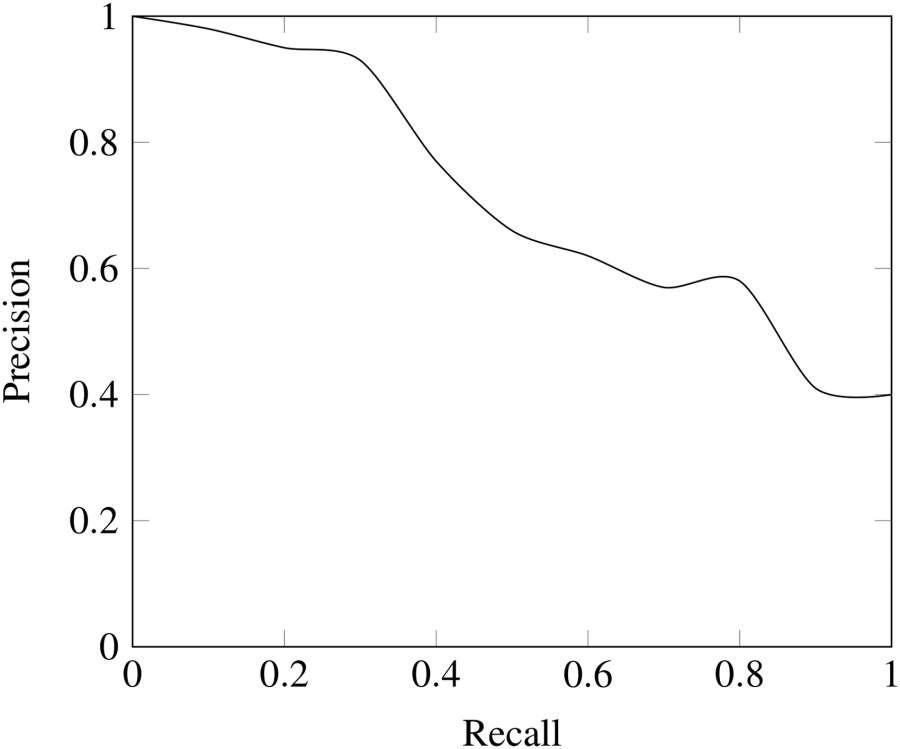






Na chłopski rozum: szukamy terrorystów w samolocie. Zakładamy, że nie ma terrorystów-mamy **accuracy** super, ale nie wykrywamy żadnych terrorystów, którzy faktycznie tam są. Druga sprawa, zakładamy, że wszyscy są terrorystami-mamy **recall** super, wyłapujemy wszystkich terrorystów, ale dużym kosztem niewinnych osób, i wreszcie trzecia sprawa, znajdujemy jednego prawdziwego terrorystę, ale pomijamy dwóch( trzech ich było w samolocie)-mamy super **precision**, ale dwóch nam uciekło. Ot i całe rozróżnienie.

The Precision-Recall Trade-off



Now, we can see that our first model which labeled all individuals as not terrorists wasn’t very useful. Although it had near-perfect accuracy, it had 0 precision and 0 recall because there were no true positives! Say we modify the model slightly, and identify a single individual correctly as a terrorist. Now, our precision will be 1.0 (no false positives) but our recall will be very low because we will still have many false negatives. If we go to the other extreme and classify all passengers as terrorists, we will have a recall of 1.0 — we’ll catch every terrorist — but our precision will be very low and we’ll detain many innocent individuals. In other words, as we increase precision we decrease recall and vice-versa.

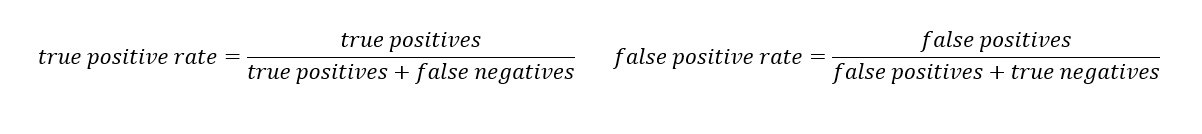
Where we want to find an optimal blend of precision and recall we can combine the two metrics using what is called the[F1 score](https://en.wikipedia.org/wiki/F1_score" \t "_blank).

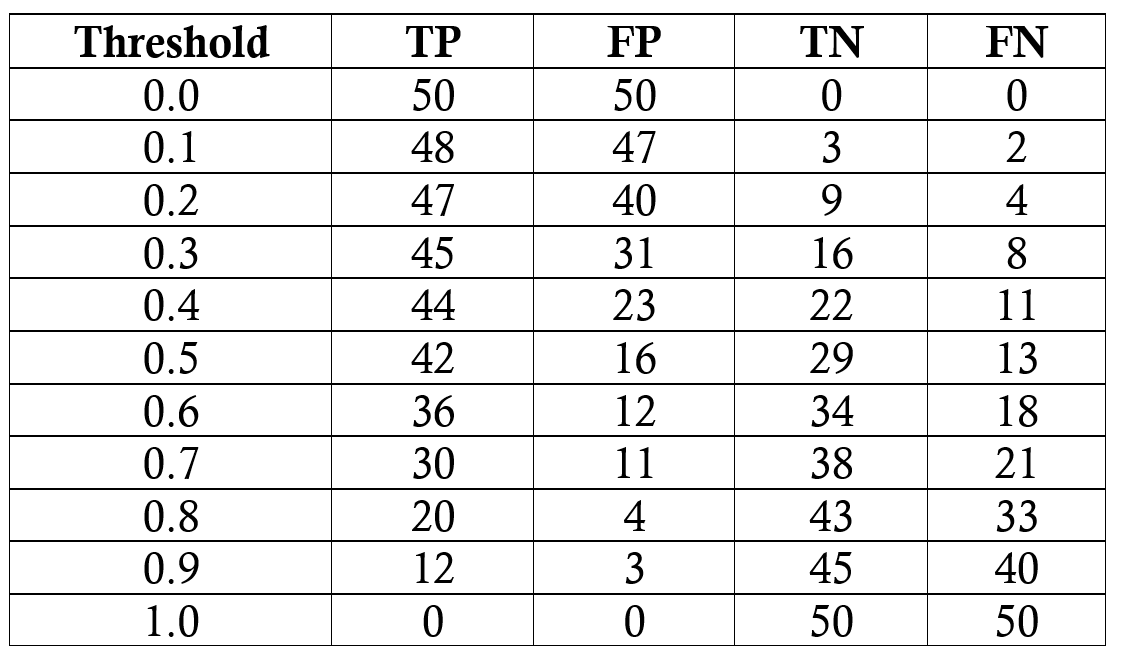


If we want to create a balanced classification model with the optimal balance of recall and precision, then we try to maximize the F1 score.

Na chłopski rozum.Threshold to balansowanie pomiędzy recall a precision. Możemy ją ustawiać dowolnie w granicach od 0 do 1, w zależności czy chcemy mieć wyższy recall, czy precision.

Aby utworzyć ROC musimy obliczyć TPR i FPR. Ważne aby nie mieszać że to jest to samo co recall i precision kiedy chcemy obliczyć F1. Otóż, aby nie było tak prosto to TPR to jest wlaśnie recall, ale FPR to nie jest precision!!!





Na chłopski rozum, jeśli threshold jest na 0 to w przykładzie z terrorystami, zakładamy że nie ma żadnego na pokładzie(accuracy super), wraz ze wzrostem threshold rośnie recall, czyli zwiększamy ilość osób uważanych za terrorystów (będących nimi czy nie to nieważne) i tym samym wyłapujemy faktycznych terrorystów. Musimy jednak uważać, bo wraz ze wzrostem threshold a tym samym i recall, maleje precision, czyli co z tego że wyłapujemy coraz więcej terrorystów kiedy musimy coraz więcej osób oczernić nimi niebędącymi.

Czyli threshold 0 oznacza, że nie ma terrorystów, a 1 oznacza, że wszyscy nimi są.

Kolor krzywej oznacza inny model. W wyborze który jest lepszy stosujemy AUC czyli obliczenia pola pod krzywą. AUC jest miarą pomiędzy 0-1. Im większe pole tym lepiej. Czarna linia to losowy klasyfikator, w tym przypadku AUC równa się 0,5.

