Task: Dog (D) vs Non-Dog (X) classification

Label = [D X D D D X X D D X]
Pred = [D D D X D X D X D D]
TP FP TP FN TP TN FP FN TP FP

"Positive" means the class in consideration. In this case it is Dog class

TP = 4 (Pred D and truth D) ⇒ True Positive
FP = 3 (Pred D but truth X) ⇒ False Positive
TN = 1 (Pred X and truth X) ⇒ True Negative
FN = 2 (Pred X " " D) ⇒ False Negative

Accuracy

$$\text{Accuracy} = \frac{TP + TN}{\text{all}} = \frac{6}{10} = 0.6$$

* Not a reliable measure by itself, be cautious.

Recall

→ Truth based measurement

→ Out of all dog labels (truths), how many we got right
Thus, the recall is the ability of a model to label a clas

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{4}{4+2} = \frac{4}{6} = 0.67$$

Low recall is an indication that the model is struggling to recognize "dog" images.

Precision

→ Prediction based measure

→ Out of all dog predictions, how many we got right!
Thus, precision is the ability of a model not to label th

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{4}{4+3} = \frac{4}{7} = 0.57$$

If the precision is low, this means the model keeps labeling non-dog images as dog images. In other words, there are too many false-positives

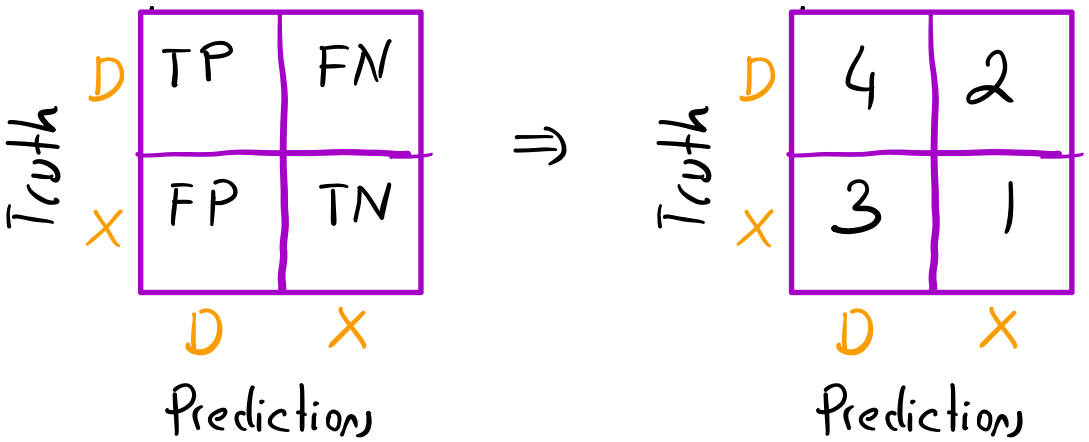
F1- Score

$$\text{F1-Score} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}}$$

This is a reliable measure putting equal weights on recall and precision.

Confusion Matrix

All relevant info can be put into a single matrix of the following form.



Rows: truths (real values)
Cols : predictions

4 dogs were identified as dog (TP)
2 dogs " " " X (FN)
3 X " " " dog (FP)
1 X " " " X (TN)

$$\text{Accuracy} = \frac{TP+TN}{all} = \frac{4+1}{10} = 0.60$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{4}{4+2} = \frac{4}{6} = 0.67$$

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{4}{4+3} = \frac{4}{7} = 0.57$$

$$F_1\text{-Score} = \frac{2}{\frac{1}{\text{Prec.}} + \frac{1}{\text{Rec}}} = 0.61$$

Dog is taken as a positive class. So all these scores are for dogs not X class

Now, let's see why we can't just rely on accuracy and why we prefer some scores over others.

Ex1 (Zombie Apocalypse)

We classify people as human or zombie and get them into a safe zone. Which score is more important for human class; recall or precision

recall: Low false-negatives, don't identify human as zombie

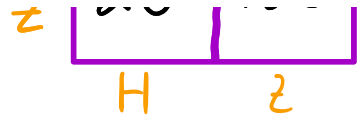
precision: low false-positive, " " zombie as human

Let's say we have the following results

1)

Truth \	H	Z
H	100	1
Z	20	150

$$\text{Acc} = \frac{100+150}{100+150+1+20} = 0.92$$



$$\text{Recall} = \frac{100}{100+1} = 0.99$$

$$\text{Precision} = \frac{100}{100+20} = 0.83$$

1 human of out 100 is identified as zombie so they stay out

17 out of 100 human prediction is wrong. This 17/100 of the time, we accept zombies into safe zone!

2)

Truth \ Prediction	H	Z
H	100	20
Z	1	150

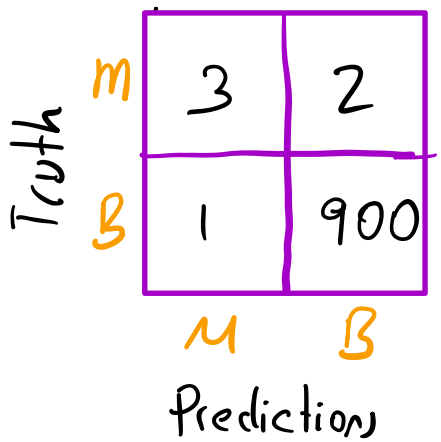
$$\text{Acc} = 0.92$$

$$\text{Recall} = \frac{100}{120} = 0.83$$

$$\text{Precision} = \frac{100}{100+1} = 0.99$$

Ex: We classify tissue samples as malignant (cancerous) or benign (non-cancerous). Malignant is the positive class. Malignant tumors are extremely rare cancer samples

1)



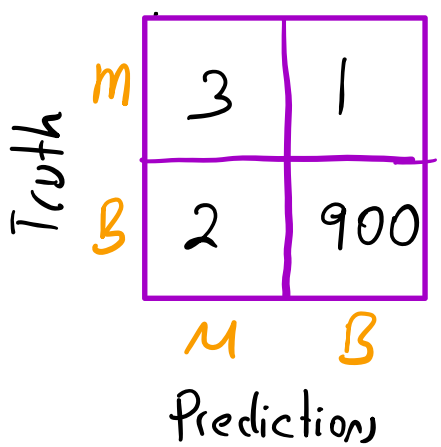
$$Acc = \frac{3+900}{3+900+2+1} = 0.99$$

$$Recall = \frac{2}{3+2} = 0.4$$

$$Precision = \frac{3}{3+1} = 0.75$$

Only 40 out of 100 malignant is correct, the rest is classified as normal.

2)



$$Acc = \frac{2+900}{2+900+1+2} = 0.99$$

$$Recall = \frac{3}{3+1} = 0.75$$

$$Precision = \frac{3}{3+2} = 0.6$$

Now 75 of 100 malignant is correct!

Recall is more important.

Multiclass Case

Ex:

Actual \ Predictions	Covid	Flu	Normal
Covid	20	25	5
Flu	7	50	12
Normal	1	4	1000

$$\text{Accuracy} = \frac{20+50+1000}{1124} = 0.95 \ !$$

Now, we can compute recall and precision for each class.

Actual \ Predictions	Covid	Others
Covid	20	30
Others	8	1066

$$\text{Recall} = \frac{20}{20+30} = 0.40$$

$$\text{Precision} = \frac{20}{20+8} = 0.71$$

$$F1 = 0.51$$

Actual \ Predicted	Flu	others
Flu	50	19
others	29	1026

$$\text{Recall} = \frac{50}{50+19} = 0.72$$

$$\text{Precision} = \frac{50}{50+29} = 0.63$$

$$F1 = 0.67$$

Actual \ Predicted	Normal	others
Normal	1000	5
others	17	102

$$\text{Recall} = \frac{1000}{1000+5} = 0.99$$

$$\text{Precision} = \frac{1000}{1000+17} = 0.98$$

$$F1 = 0.98$$

truth based

prediction based

	Recall	Precision
Covid	40 %	71 %
Flu	72 %	63 %
Normal	99 %	98 %

% 40 → out of 100 covid cases, we catch only 40

% 71 → out of 100 covid predictions, we get 71 right