### Precision

Precision focuses on the **predicted** "positive" values in your dataset. By optimizing based on precision values, you are determining if you are doing a good job of predicting the positive values, as compared to predicting negative values as positive.

|  |  |  |
| --- | --- | --- |
|  | Spam (positive) | Not Spam (negative) |
| Spam (positive) | True Positive(TP) | False Positive(FP) |
| Not Spam (negative) | False Negative(FN) | True Negative(TN) |

### Medical model

So let's define the precision metric as follows. Here's the confusion matrix of the medical model, and we've added a red X in the spot that we really want to avoid, which is the false negatives.

precision is the answer to the question, "*Out of all the points predicted to be positive, how many of them were actually positive?*" In this case, the question translates to, out of all the patients that we diagnosed as sick, how many were actually sick? So precision is this column because this column is the sick patients that we diagnose as sick. So it is 1,000 that were correct, divided by 1,800, which is the total number of patients diagnosed as sick. This number is 55.7 percent. It's not a high number because this is not a very precise model. But, again, this is okay because what we're doing is avoiding this red X.

### Spam email model

What is this precision? Now, we know that in this model, precision is very important because the red X that we're avoiding is in this column. The red X is the non-spam email that was accidentally sent to the spam folder. So those 30 errors are really bad and we want to avoid them. So, again, precision says, '*'Out of all the emails sent to the spam folder, how many of them were actually spam?*" So we have 100 which are correct, divided by 130 which is all the ones we've sent to the spam folder. This number is 76.9 percent, which is higher. This is better since this model needs high precision, so the number better be big.

#### **Action Required**

Now, let's do an exercise. Let's go to a linear model over here. What is the precision of this linear model?



### **Precision Quiz**

What is the precision of the linear model above? Please write the number as a decimal, like **0.45** or as a fraction, like **3/5**.

Answer: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Given the confusion matrix from the previous question:

True Positives (TP) = 3 (blue points above the line)

False Positives (FP) = 1 (red points above the line)

Precision Formula:

Precision = TP / (TP + FP) = 3 / (3 + 1) = 3/4 or 0.75

Final Answer:

0.75 (or 3/4)

### **Quiz Question**

So precision will be the answer to the question?

1. "Out of all the points predicted to be positive, how many of them were actually positive?"
2. "Out of all the points predicted to be positive, how many of them were actually negative?"