### Recall

Recall focuses on the **actual** "positive" values in your dataset. By optimizing based on recall values, you are determining if you are doing a good job of predicting the positive values **without** regard to how you are doing on the **actual** negative values. If you want to perform something similar to recall the **actual** 'negative' values, this is called specificity (TN / (TN + FP)).

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

​

### Medical model

Let's look at the second metric called **recall**. Recall is the answer to the following question, "*out of the points that are labeled positive, how many of them were correctly predicted as positive?*"

So in the medical model, a recall is the answer to the following question, "*out of the sick patients, how many did we correctly diagnose as sick?*" Remember:

* **Precision** was, "*out of the patients were diagnosed as sick, how many were actually sick?*"
* **Recall** is the opposite, "*out of the patients that are sick, how many did we correctly diagnose as sick?*"

It can be seen as a reach of the algorithm. How many of the positive points did I manage to catch? As we can see, this row catches a critical case labeled by an X, so we can see that recall is important in the medical model. Now to calculate recall, we do the following; from the 1,200 sick patients, how many did we diagnose correctly? That's 1,000 divided by 1,200, which is 83.3 percent. This model better have a higher recall because we're trying to correctly catch as many of the sick people as possible.

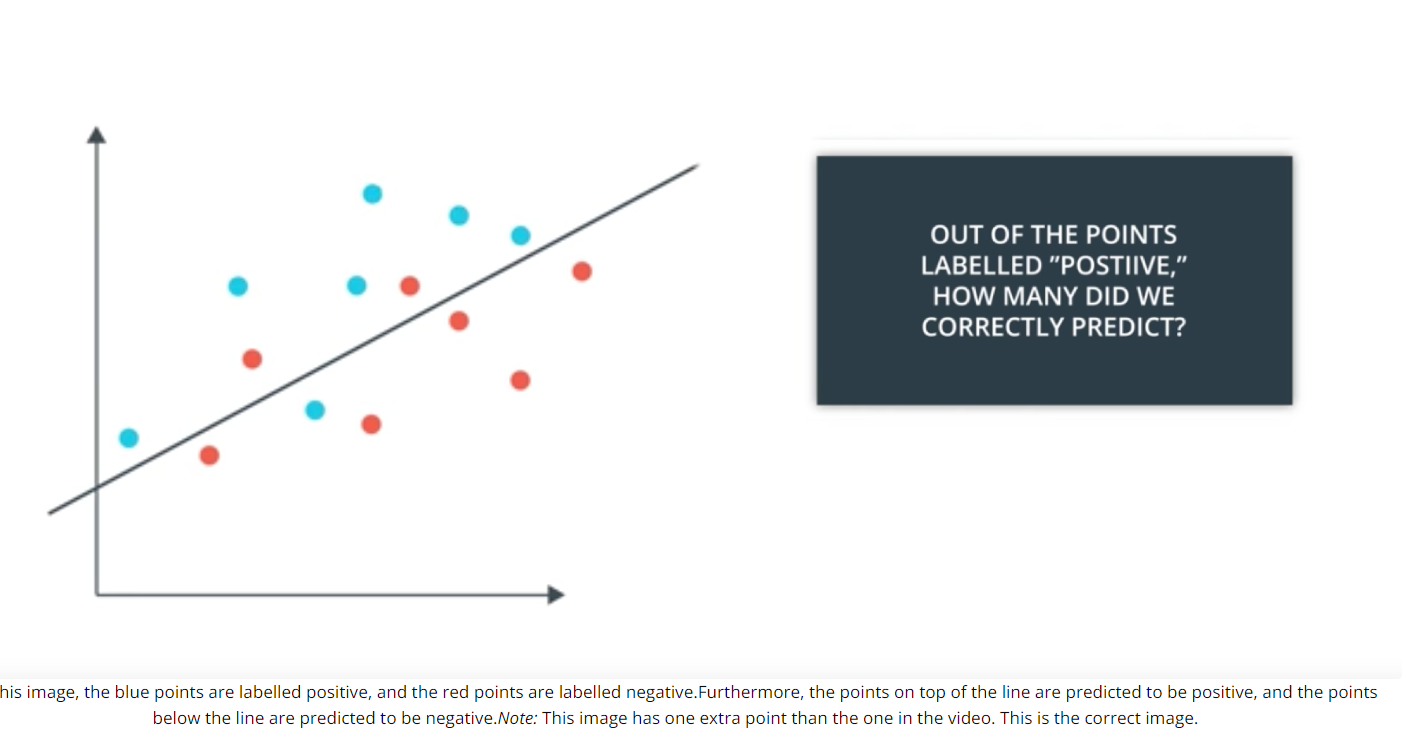
### Spam email model

Recall is calculated as follows; from the spam emails, how many of them do we correctly send to the spam folder? You can see the row with 100 correctly sent to the spam folder divided by 270 total spam emails. It's a low number. It's 37 percent. But remember, we are worried about avoiding this X, since we don't mind if we don't catch all the spam emails as long as the ones we caught are spam.

It's okay that this model has a low recall.

### Action Required

Let's do a similar exercise as before. In this linear model over here, what is the recall?



### 

### **Recall Quiz**

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

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

True Positives (TP) = Blue points above the line (correctly predicted positive)

False Negatives (FN) = Blue points below the line (incorrectly predicted negative)

Recall Formula:

Recall = TP / (TP + FN)

Example Calculation (assuming counts from previous questions):

If there are 3 True Positives and 2 False Negatives:

Recall = 3 / (3 + 2) = 3/5 or 0.60

Final Answer:

0.60 (or 3/5)

### **Quiz Question**

Recall is the answer to the following question:

1. "Out of the points that are labeled positive, how many of them were correctly predicted as negative?"
2. "Out of the points that are labeled positive, how many of them were correctly predicted as positive?"