Suppose you have 100 documents and a search query of “ford”. Of the 100 documents, suppose 80 are related/relevant to the term “ford”, and the other 20 are not relevant to “ford”.

1 Scenario 1:

Now suppose your search algorithm returns 60 result documents where 45 docs are in fact relevant and 15 result docs are not relevant .
There were 35 relevant documents which were not retrieved i.e out of 80 relevant documents, only 45 are retrieved(80 - 45 = 35).
There are 5 irrelevant documents which are correctly omitted i.e out of 20 irrelevant documents, 15 documents were incorrectly retrieved as relevant (so 20 - 15 = 5).

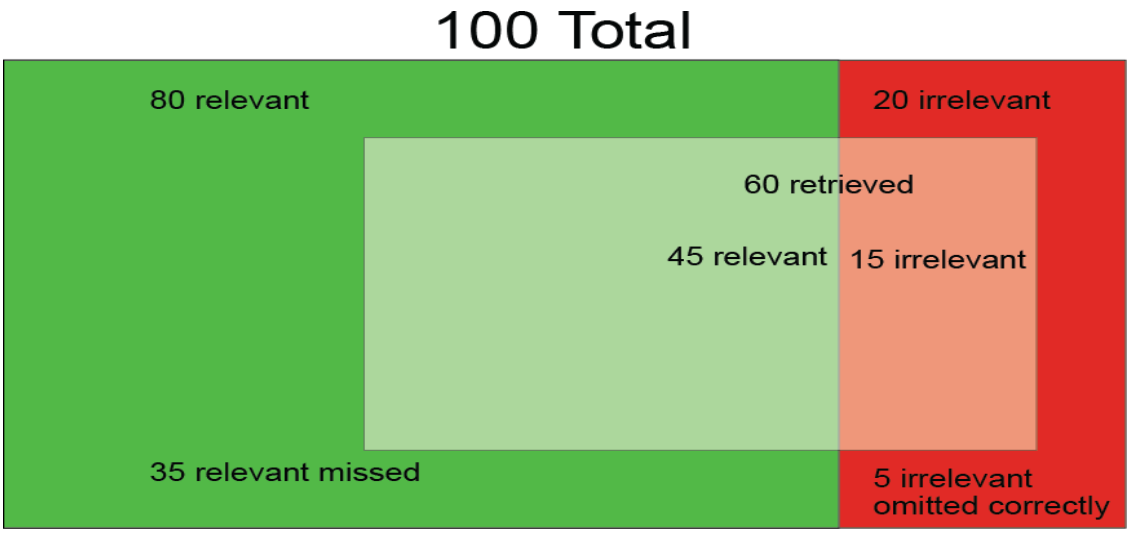

Figure 1

From the above, we can come to an understanding that,
True Positives (TP) = 45 (45 documents are relevant and correctly retrieved)
False Positives (FP) = 15 (15 documents are irrelevant, but incorrectly retrieved)
False Negatives (FN) = 35 (35 documents are relevant, but incorrectly omitted)
True Negatives (TN) = 5 (5 documents are irrelevant and correctly omitted)

The **precision** is the fraction/percentage of retrieved docs that are relevant.

$$precision = \frac{retrieved\ relevant}{total\ relevant}$$

The **recall** is the fraction/percentage of relevant docs that were retrieved.

$$recall = \frac{retrieved\ relevant}{retrieved\ relevant + retrieved\ irrelevant}$$

$$precision = \frac{45\ retrieved\ relevant}{45\ retrieved\ relevant + 15\ retrieved\ irrelevant}$$

$$= 0.75$$

$$recall = \frac{45\ retrieved\ relevant}{45\ retrieved\ relevant + 35\ not\ retrieved\ relevant}$$

$$= 0.56$$

In short, both precision and recall in information retrieval are measures of goodness that are tied to the notion of relevance. Precision and recall in machine learning binary classification are very different.

In the above scenario, if the number of retrieved documents were less than the number of relevant documents, the recall was 56% and precision was 75%.

Now let us assume the number of retrieved documents are greater than the number of relevant documents.

2 Scenario 2:

Now suppose your search algorithm returns 90 result documents where 80 docs are in fact relevant and 10 result docs are not relevant.

There are no relevant documents which were not retrieved i.e all 80 relevant docs are retrieved.

There are 10 irrelevant documents which are correctly omitted i.e out of 20 irrelevant documents, 10 documents were incorrectly retrieved as relevant (so 20 - 10 = 10).

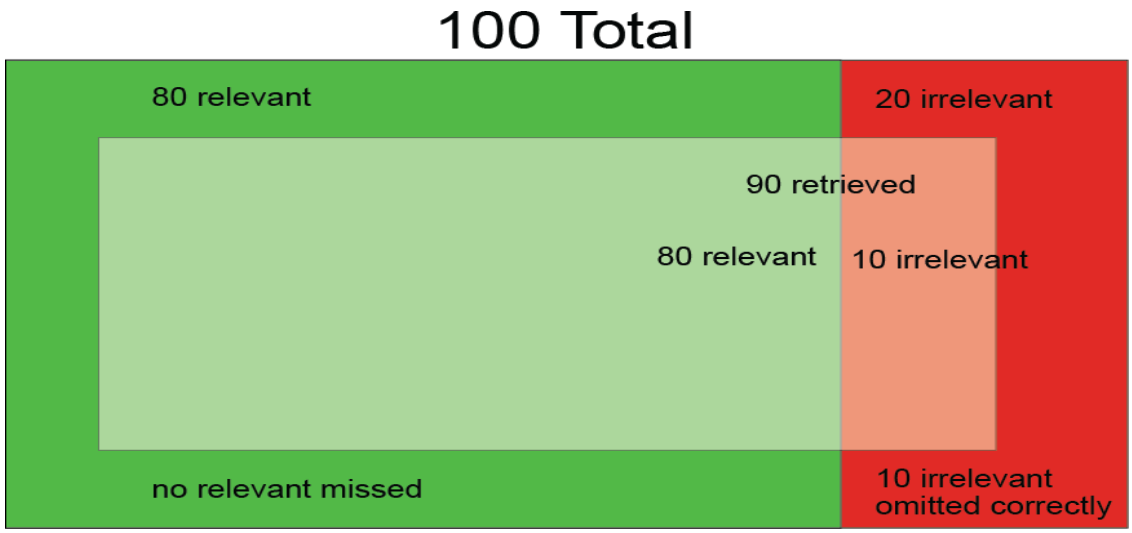

Figure 2

From the above, we can come to an understanding that,
True Positives (TP) = 80 (80 documents are relevant and correctly retrieved)
False Positives (FP) = 10 (10 documents are irrelevant, but incorrectly retrieved)
False Negatives (FN) = 0 (No relevant documents are incorrectly omitted)
True Negatives (TN) = 10 (10 documents are irrelevant and correctly omitted)

$$precision = \frac{80\ retrieved\ relevant}{80\ retrieved\ relevant + 10\ retrieved\ irrelevant}$$

$$= 0.88$$

$$recall = \frac{80\ retrieved\ relevant}{80\ retrieved\ relevant + 0\ not\ retrieved\ relevant}$$

$$= 0.1$$

In the second scenario, if the number of retrieved documents were greater than the number of relevant documents, the recall was 100% and precision was 88%.

From this, we can gather that one can reach 100% recall. (As all the 80 relevant documents were retrieved correctly without any relevant document being omitted incorrectly.)

3 Scenario 3:

Now suppose your search algorithm returns 60 result documents where 60 docs are in fact relevant and 0 result docs are not relevant .
There are 20 relevant documents which were not retrieved i.e out of 80 relevant docs, only 60 are retrieved (80 - 60 = 20).
There are 20 irrelevant documents which are correctly omitted i.e no irrelevant document was incorrectly retrieved as relevant .

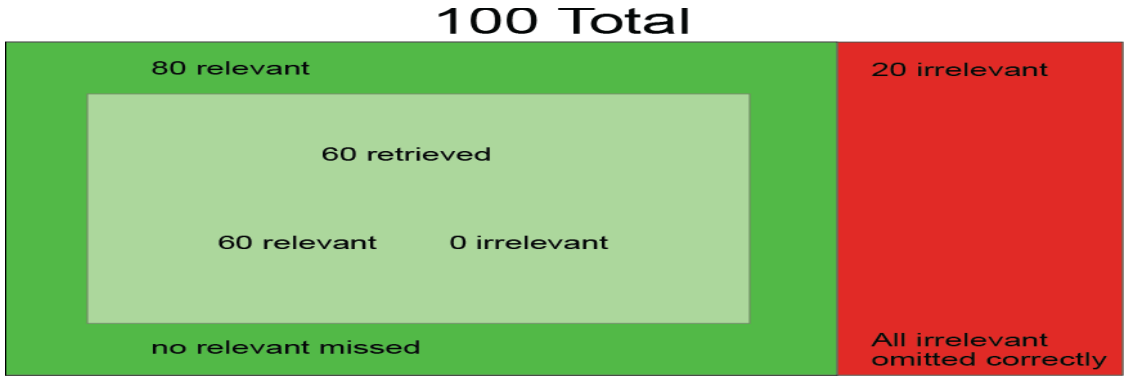

Figure 3

From the above, we can come to an understanding that,
True Positives (TP) = 60 (80 documents are relevant and correctly retrieved)
False Positives (FP) = 0 (No documents are irrelevant, but incorrectly retrieved)
False Negatives (FN) = 20 (20 relevant documents are incorrectly omitted)
True Negatives (TN) = 20 (10 documents are irrelevant and correctly omitted)

$$precision = \frac{80\ retrieved\ relevant}{80\ retrieved\ relevant + 0\ retrieved\ irrelevant}$$

$$= 0.1$$

$$recall = \frac{80\ retrieved\ relevant}{80\ retrieved\ relevant + 20\ not\ retrieved\ relevant}$$

$$= 0.8$$

In the third scenario, if the number of retrieved documents were less than the number of relevant documents, and all the retrieved documents were relevant the precision was 100% and recall was 80%.

4 Scenario 4:

Now suppose your search algorithm returns 80 result documents where 80 docs are in fact relevant and 0 result docs are not relevant .
There are no relevant documents which were not retrieved i.e all 80 relevant docs are retrieved.
There are 20 irrelevant documents which are correctly omitted i.e no irrelevant document was incorrectly retrieved as relevant .

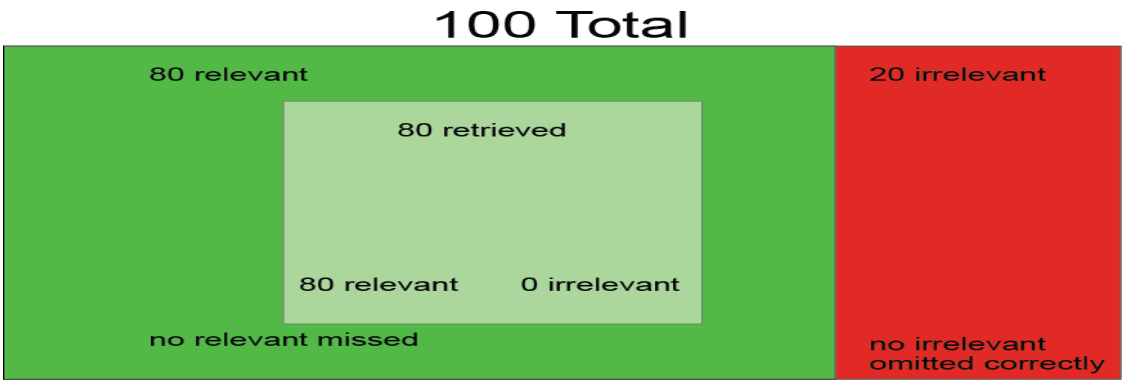

Figure 4

From the above, we can come to an understanding that,
True Positives (TP) = 80 (80 documents are relevant and correctly retrieved)
False Positives (FP) = 0 (No documents are irrelevant, but incorrectly retrieved)
False Negatives (FN) = 0 (No relevant documents are incorrectly omitted)
True Negatives (TN) = 20 (10 documents are irrelevant and correctly omitted)

$$precision = \frac{80\ retrieved\ relevant}{80\ retrieved\ relevant + 0\ retrieved\ irrelevant}$$

$$= 0.1$$

$$recall = \frac{80\ retrieved\ relevant}{80\ retrieved\ relevant + 0\ not\ retrieved\ relevant}$$

$$= 0.1$$

In the fourth scenario, if the number of retrieved documents were greater than the number of relevant documents, the recall was 100% and precision was 100%.

From this, we can gather that one can reach 100% recall and 100% precision. (As all 80 relevant docs are retrieved and no irrelevant doc is retrieved).

The precision and recall are inversely related.

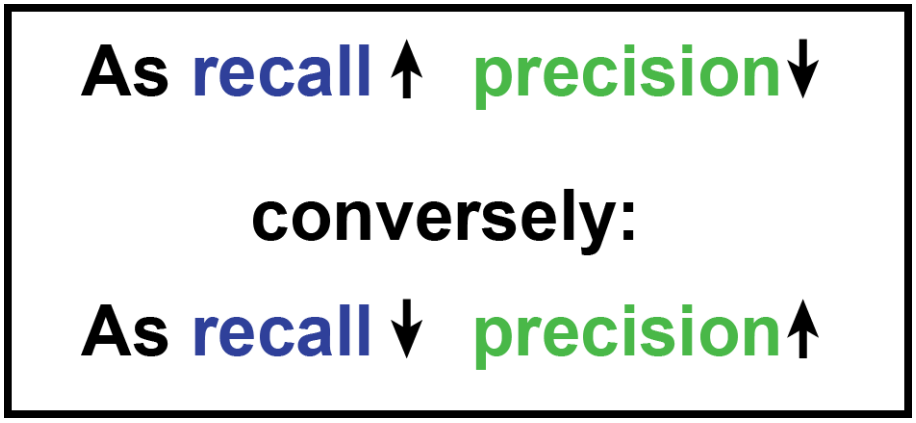

Figure 5

Here is an exercise for you to work on:

Let's say we are solving a classification problem where we are predicting whether a person is having cancer or not.

Let's give a label to our target variable:
YES: When a person is having cancer
NO: When a person is NOT having cancer.

Now that we have identified the problem, the confusion matrix, is a table with two dimensions ("Actual" and "Predicted"), and sets of "classes" in both dimensions. Our Actual classifications are columns and Predicted ones are Rows.

n=165	**Predicted: NO**	**Predicted: YES**	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Figure 6

Perform the calculations required to arrive at the following values for Accuracy, Precision and Recall.

Accuracy = 0.91

Precision = 0.91

Recall = 0.95

Check your understanding:

- There are 15.6 million results to a query from Google. The relevant links to the question are 2 million. Assuming there are also about 6 million more results that were relevant but weren't returned by Google, for such a system, what is the precision and recall?

    a) precision = 0.13 recall = 0.25

    b) precision = 0.25 recall = 0.13

    c) precision = 0.66 recall = 0.59

    d) precision = 0.19 recall = 0.78

- If someone asked you to list the names of 5 presents you got last Christmas but you couldn't exactly remember the 5 names, so you randomly guessed seven times. Out of the 7 names you remembered, 5 was recalled correctly.The other 2 were gifts you received on your birthday. Even though you got a 100% recall (5/5), what would be your precision?

    a) precision = 82%

    b) precision = 68.9%

    c) precision = 71.4%

    d) precision = 65.7%