The Precision/Recall Tradeoff

Intro
When applying machine learning and more specifically when using classifiers, it is important to think about the precision/recall tradeoff in regards to what goal your system has.
Ideally you would probably strive to make your system high in both precision and recall, but at some point, if you want to get the last percentages, you will have to face reality and choose between very high precision or very high recall, or at least have made a conscious decision about what you want to focus.

Discussion
To clarify, when you want your precision high you want to be sure you have as few false positives as possible, which also comes with a higher number of false negatives. On the other hand when you want your recall high you want to have as few false negatives as possible, which then inevitably will raise the number of false positives you get.
Thus, in examples of a system that detects cancer cells in blood samples, you would probably want your system to have a very high recall, and then have a human making a more precise sorting, getting rid of potential false positives. A system that would require high precision could be one where you would get so much data that the ones you have to look at, you want to make sure they are valid.

When getting to the math you'll have these two equations to determine the percentage of precision and recall:

$$Precision = \frac{TruePositive}{TruePrositive + FalsePositive}$$

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

Changing the percentage to favor either precision or recall can be done by raising or lowering the threshold.

From the book Hands-on Machine Learning with Scikit-Learn & TensorFlow, figure 3-4 shows how altering the threshold is connected to get you a higher precision or recall depending on which way you move the threshold.
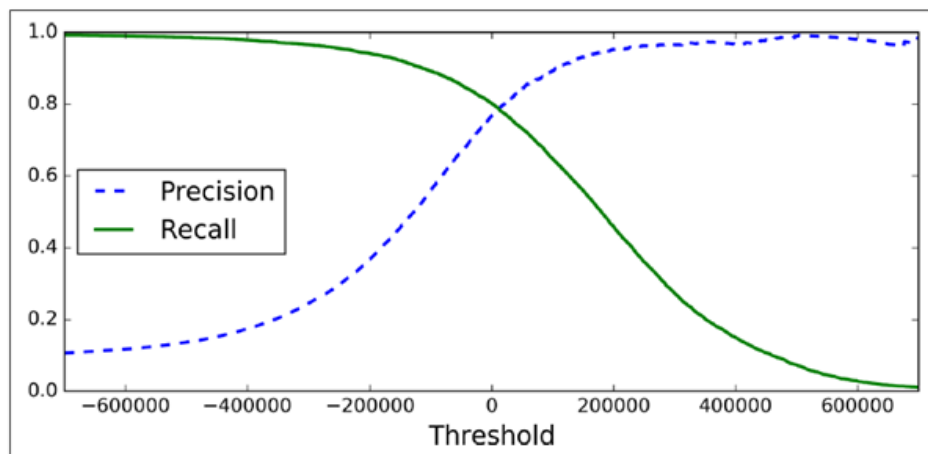

Figure 3-4. Precision and recall versus the decision threshold

This graph is created using the MNIST dataset containing 70000, samples of handwritten numbers 1-9.

Looking at this graph it's also important to notice that even though you might want a high recall or high precision, you obviously need to think of the question; at what cost? Having 99% in precision doesn't make much sense when your tradeoff is going to be 15% recall, here you'd be much better served going for 90%-95% precision and then getting 70%-60% recall as the tradeoff

References
Hands-on Machine Learning with Scikit-Learn & TensorFlow by Aurélien Géron, 1st edition.