Trading Off Precision and Recall

For many applications we want to control the trade-off between precision and recall

- Logistic regression: $0 \leq h_\theta(x) \leq 1$
- Predict 1 if $h_\theta(x) \geq 0.5$
- Predict 0 if $h_\theta(x) < 0.5$

Suppose we want to predict $y = 1$ (cancer) only if very confident. One way to do this would be to modify the algorithm, so that instead of setting this threshold at 0.5, we might instead say that we will predict that $y = 1$ only if $h_\theta(x) \geq 0.7$ and so we end up with a classifier that has **higher precision**, but in contrast this classifier will have **lower recall**.

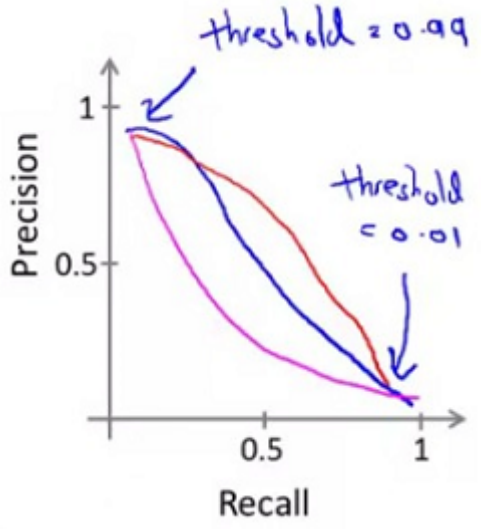
- Logistic regression: $0 \leq h_\theta(x) \leq 1$
- Predict 1 if $h_\theta(x) \geq 0.7$
- Predict 0 if $h_\theta(x) < 0.7$
- **Higher precision, lower recall**
  - Lower Recall, because now we're going to predict $y = 1$ on a smaller number of patients (risk of false negatives).

Another example: Suppose we want to avoid missing too many cases of cancer (avoid false negatives). In particular, if a patient actually has cancer, but we fail to tell them that they have cancer then that can be really bad. Because if we tell a patient that they don't have cancer, then they're not going to go for treatment. In this case, rather than setting higher probability threshold, we might instead take the threshold probability value and instead set it to a lower value (i.e. 30% chance they have cancer):

- Predict 1 if $h_\theta(x) \geq 0.3$
- Predict 0 if $h_\theta(x) < 0.3$
- **Higher recall, lower precision**
  - Higher Recall, because we're going to be correctly flagging a higher fraction of all of the patients that actually do have cancer.
  - With lower Precision, we have more risk of false positives, because we're less discriminating in deciding what means the person has cancer.

More generally. Predict 1 if $h_\theta(x) \geq$ threshold.

- For most classifiers there is going to be a trade off between precision and recall, we can show this graphically by plotting precision vs. recall



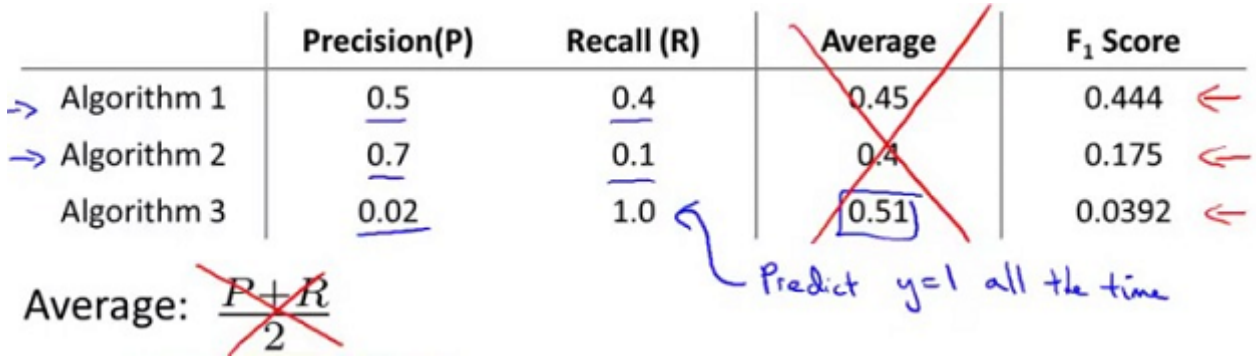
This curve can take many different shapes depending on classifier details

F_1 Score (F Score)

How to compare precision/recall numbers?

Concretely, suppose we have three different learning algorithms. So actually, maybe these are three different learning algorithms, maybe these are the same algorithm but just with different values for the threshold. How do we decide which of these algorithms is best?

We spoke previously about the importance of a single real number evaluation metric, we have actually lost that, we now have two real numbers (By switching to precision/recall we have two numbers). how can we get a single real number evaluation metric?

- One option is the average
  - Average: $(P + R) / 2$
  - This is not such a good solution
    - Means if we have a classifier which predicts $y = 1$ all the time we get a high recall and low precision. Similarly, if we predict y rarely get high precision and low recall.
    - So the average of precision and recall as not a particularly good way to evaluate our learning algorithm.
- 

- One way for combining precision and recall is using F Score
  - Fscore is like taking the average of precision and recall giving a higher weight to the lower value
  - If P = 0 or R = 0 --> Fscore = 0
  - If P = 1 and R = 1 --> Fscore = 1
  - The remaining values lie between 0 and 1

Threshold offers a way to control trade-off between precision and recall. Concretely, Fscore gives a single real number evaluation metric.

Video Question: You have trained a logistic regression classifier and plan to make predictions according to:

- Predict $y = 1$ if $h_\theta(x) \geq \text{threshold}$
- Predict $y = 0$ if $h_\theta(x) < \text{threshold}$

For different values of the threshold parameter, you get different values of precision (P) and recall (R). Which of of the following would be a reasonable way to pick the value to use for the threshold?

- Measure precision (P) and recall (R) on the **test set** and choose the value of threshold which maximizes $\frac{P+R}{2}$
- Measure precision (P) and recall (R) on the **test set** and choose the value of threshold which maximizes $2\frac{PR}{P+R}$
- Measure precision (P) and recall (R) on the **cross validation set** and choose the value of threshold which maximizes $\frac{P+R}{2}$

> Measure precision (P) and recall (R) on the **cross validation set** and choose the value of threshold which maximizes $2\frac{PR}{P+R}$