DAY7 Ben, Bram, Kaiyu, Qingyuan, Yuzhe

* SGDClassifier is a linear classifier which implements regularized linear models with stochastic gradient descent (SGD) learning.
* Precision: accuracy of the positive predictions. (precision = TP/(TP+FP))
* Recall: also called sensitivity or true positive rate, this is the ratio of positive instances that are correctly detected by the classifier. (recall = TP/(TP+FN))
* 

* Precision/recall tradeoff
* 
* How SGDClassifier makes its classification decisions. It computes a score based on a decision function, and if that score is greater than a threshold, it assigns the instance to the positive class, or else it assigns it to the negative class.
1. The decision threshold is positioned at the central arrow: you will find 4 true positive (actual 5s) on the right of that threshold, and one false positive (actually a 6).

Precision = 80%

But out of 6 actual 5s, the classifier only detects 4.

Recall = 67%

1. Raise the threshold, move it to the right.

The false positive (the 6) becomes a true negative.

Precision = 100%

Bit one true positive becomes a false negative.

Recall = 50%

……

* You can call decision\_function() method, which returns a score for each instance, and then make predictions based on those scores using any threshold you want:



* How can you decide which threshold to use?
1. Get the scores of all instances in the training set using the cross\_val\_predict() function again:



1. Compute precision and recall for all possible thresholds using the precision\_recall\_curve() function:



1. Plot precision and recall as functions of the threshold value using Matplotlib:





Other methods:

Plot precision directly against recall:

