Quiz 3

Solutions

Please write all answers on these pages.

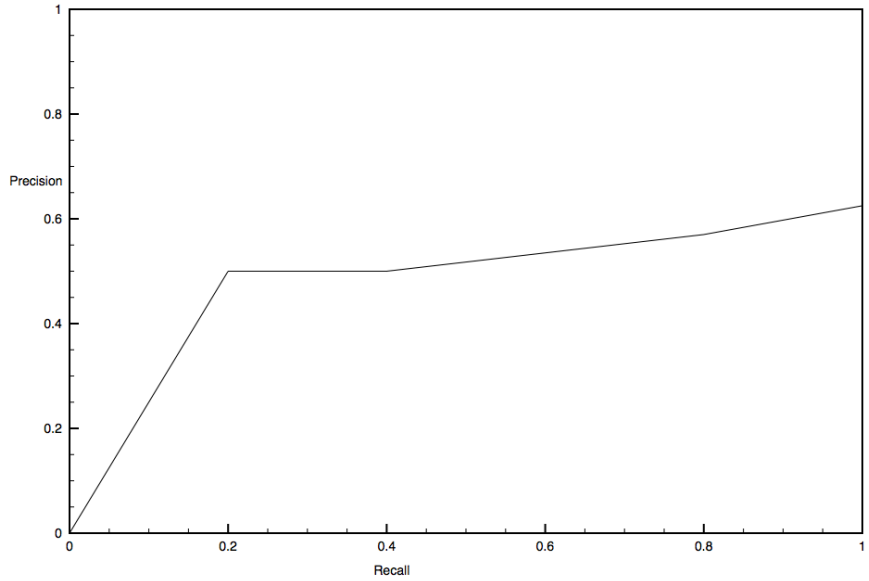
1. Consider the table below, which reports, for four instances, the actual class of each instance and the score given by a classifier to that instance.

Instance	Actual Class	Score
1	Negative	5
2	Positive	3
3	Positive	2
4	Negative	1
5	Positive	-1
6	Positive	-2
7	Negative	-3
8	Positive	-6

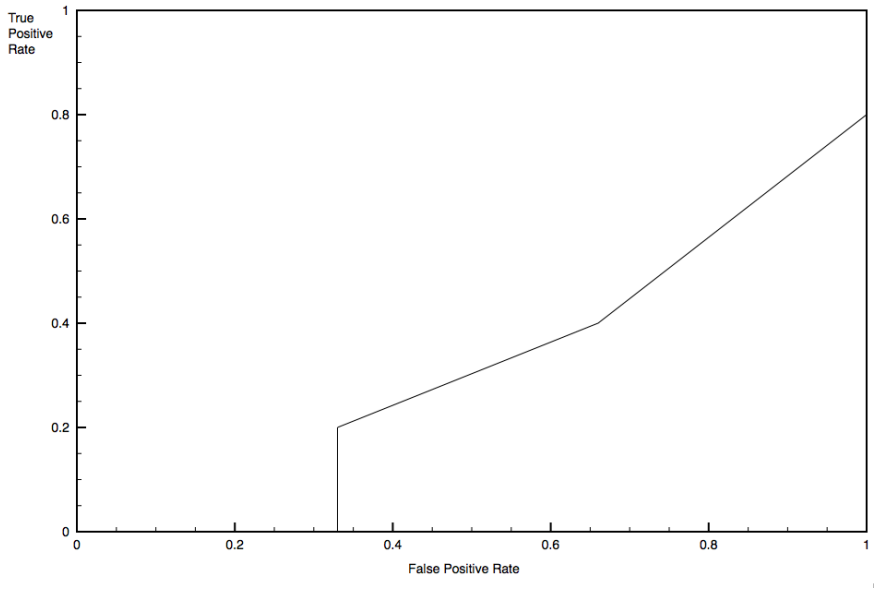
(a) For each value of the threshold in the table below, fill in the accuracy, precision, and recall of the classifier. (If the score is equal to or greater than the threshold, the classifier assigns "positive", else it assigns "negative".) *Note that when there are no true positives or false positives, precision = 1.*

Threshold	Accuracy	Precision	Recall
5	2/8	0	0
3	3/8	1/2	1/5
1	3/8	1/2	2/5
-3	4/8	4/7	4/5
-6	5/8	5/8	1

(b) Sketch a precision-recall curve with these five thresholds. (Be sure to label the axes.)



(c) Sketch a ROC curve with these five thresholds. (Be sure to label the axes.)



2. Suppose you have a training set with 1,000 labeled instances and a test set with 400 labeled instances. You want to train and test a neural network using this dataset, but you also want to decide whether 10, 20, or 30 hidden units will give the best generalization

performance (keeping all other hyperparmeters fixed). Explain the main steps you would take to do this using 10-fold cross-validation on the training set.

1. Divide training set into 10 equal segments, S_i.

2. For j in {10, 20, 30}

    For i = 1 to 10

        Select S_i to be the validation set

        Train the neural network with j hidden units

        Test the classifier on S_i to obtain accuracy A_i

    Compute Average A'_j = Average over the A_i's

3. Choose value of j with highest A'_j

4. Train a neural network with j hidden units on the entire training set.

5. Test that neural network on the test set. Report this performance.

3. Label the following possible sources of classifier error as *bias*, *variance*, or *noise*.

(a) Using a perceptron to classify a non-linearly separable dataset.

Bias

(b) Using a training set whose values are specific to a particular time of year.

Variance

(c) Using a polynomial kernel with low degree when training an SVM.

Bias

(d) Using a training set with values that were entered incorrectly.

Noise

(e) Using a training set that is not "iid" (independent and identically distributed) whereas the test set is iid.

Variance