Lab 5b Cross Validation Code

Friday, July 03, 2009
10:59 AM

CROSS VALIDATION

In yesterday's lecture, we covered k-fold cross-validation.
You'll need some of this code and information to calculate your accuracy rate on your classifiers.

EXAMPLE

Let's say we have 10-fold cross validation...

1. Divide test set into 10 random subsets.
2. 1 test set is tested using the classifier trained on the remaining 9.
3. We then do test/train on all of the other sets and average the percentages.

To achieve the first step (divide our training set into k disjoint subsets), use the function **crossvalind.m** (posted in the **Utilities**)

> INDICES = CROSSVALIND('Kfold',N,K) returns randomly generated indices
> for a K-fold cross-validation of N observations. INDICES contains equal
> (or approximately equal) proportions of the integers 1 through K that
> define a partition of the N observations into K disjoint subsets.
>
> You can type **help crossvalind** to look at all the other options.

Here is an outline of how to perform cross-validation on a classifier:

```
% cross_validation
k = 10;                          % how many folds do you want?
N =  size(features,1) ;          % this is the total number of observations or rows that we have
indices = crossvalind('Kfold',N,k)  % divide test set into 10 random subsets
for i = 1:10

     % SEGMENT DATA INTO FOLDS
     disp(['fold: ' num2str(i)])
     test = (indices == i) ;     % which points are in the test set
    train = ~test;               % all points that are NOT in the test set

     % SCALE
     [trainingFeatures,mf,sf]=scale(features(train,:));

     % BUILD NEW MODEL
      model = knn(numFeatures,1,1,trainingFeatures,labels(train,:));

     % EVALUATE WITH TEST DATA
     model_output = knnfwd(model ,features(test,:))

     % COUNT ERRORS
     errors(i) = mean ( model_output ~= labels(test,:) )

end
disp(['cross validation error: ' num2str(mean(errors))])
```