Evaluating and using ML classifiers: model selection

Chapter 2

Evaluating classifier performance

The simplest evaluation protocol:

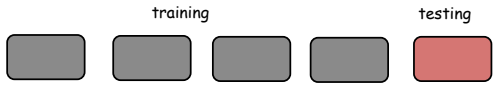
- Divide your labeled data into a training set and test set.
- Train a classifier on the training set
- Classify the examples in the test set, and measure accuracy

This tells you how well the classifier is performing on a given setting of classifier parameters.

Cross-validation

Cross validation:

- Randomly partition the data into k parts ("folds").
- Set one fold aside for testing and train a model on the remaining k-1 folds and evaluate it on the test fold.
- Repeat until each fold has been used in testing



The reported accuracy is the average over the accuracies for each fold

This tells you how well the classifier is performing on a given setting of classifier parameters.

Model selection

You have been tasked with deploying a classifier for a given task and you have some labeled data to work with.

You would like to compare several classification methods and choose the best one. Each classifier has one or more hyperparameters (e.g. SVM soft margin constant and kernel parameter).

Approach:

For each classifier compare the accuracy of the best parameter setting (estimated using cross-validation or a test set)

Model selection

You would like to compare several classification methods and choose the best one. Each classifier has one or more hyperparameters (e.g. SVM soft margin constant and kernel parameter).

Approach:

For each classifier compare the accuracy of the best parameter setting (estimated using cross-validation or a test set)

So, assuming we are comparing two classifiers, this means we are making the following comparison:

$$\max(s_1,\ldots,s_k) \quad \text{vs} \quad \max(t_1,\ldots,t_m)$$

In computing the maximum we are using information about the test set labels!

5

Two ways of doing cross validation

External cross validation:

- Perform cross validation across various settings of classifier parameters and report the best result you got

Internal cross validation (nested CV):

- For each fold, perform cross-validation on the training data, and train a classifier on the best set of parameters for that fold

6

Internal vs External cross-validation estimates

Data Set	External	Internal	Bias
banana	10.355 ± 0.146	10.495 ± 0.158	0.140 ± 0.035
breast cancer	26.280 ± 0.232	27.470 ± 0.250	1.190 ± 0.135
diabetis	22.891 ± 0.127	23.056 ± 0.134	0.165 ± 0.050
flare solar	34.518 ± 0.172	34.707 ± 0.179	0.189 ± 0.051
german	23.999 ± 0.117	24.217 ± 0.125	0.219 ± 0.045
heart	16.335 ± 0.214	16.571 ± 0.220	0.235 ± 0.073
image	3.081 ± 0.102	3.173 ± 0.112	0.092 ± 0.035
ringnorm	1.567 ± 0.058	1.607 ± 0.057	0.040 ± 0.014
splice	10.930 ± 0.219	11.170 ± 0.280	0.240 ± 0.152
thyroid	3.743 ± 0.137	4.279 ± 0.152	0.536 ± 0.073
titanic	22.167 ± 0.434	22.487 ± 0.442	0.320 ± 0.077
twonorm	2.480 ± 0.067	2.502 ± 0.070	0.022 ± 0.021
waveform	9.613 ± 0.168	9.815 ± 0.183	0.203 ± 0.064

Table 8: Error rate estimates for kernel ridge regression over thirteen benchmark data sets, for model selection schemes that are internal and external to the cross-validation process. The results for each approach and the relative bias are presented in the form of the mean error rate over for 100 realisations of each data set (20 in the case of the image and splice data sets), along with the associated standard error.

Table from
On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation
Gavin C. Cawley, Nicola L.C. Talbot, JMLR 11:2079–2107, 2010.
http://jmlr.org/papers/v11/cawley10a.html

7

Internal cross-validation

Notice that each train/test fold may get different parameter settings!

That's fine (and proper)

This results in a "parameterless" algorithm that internally sets parameters for each data set it gets

8

What to do for the system you are deploying

Use external cross-validation to determine good parameters

Train your model on ALL the data.

Provide your "customer" with the results of internal-cross validation as estimates of future performance.

9

What about this?

Do a cross-validation study to set parameters

Do another cross-validation study, using the best parameters, to estimate future accuracy

- How will this relate to the "true" future accuracy?
- Likely to be an overestimate

What about:

1. Do a proper internal cross-validation experiment
2. Improve your algorithm; goto 1

10

Over-estimates in algorithm development

Do a cross-validation study to set parameters

Do another cross-validation study, using the best parameters, to estimate future accuracy

- How will this relate to the "true" future accuracy?
- Likely to be an overestimate

What about:

1. Do a proper internal cross-validation experiment
2. Improve your algorithm; goto 1

(Machine Learning's dirty secret!)

11

Training/validation/test

If you have a lot of data you can substitute internal cross-validation with use of a training/validation/test set.

For each parameter setting, train on the training set, and choose the parameter setting that gives best performance on the validation set. Retrain using those parameters on the training + validation sets and report accuracy on the test set.

12

Correct classifier evaluation

When running experiments consider the following question:

On each fold of cross-validation, did I ever access in any way the label of a test case?

Any preprocessing done over entire data set (feature selection, parameter tuning, threshold selection) must not use labels

13

Using repository data for classifier evaluation

Pros:

- Very easy to implement
- Data from real applications
- Facilitates replication and comparison of results

Cons:

- Not representative of the data mining process which involves many steps other than classification.
- Community experiment/multiplicity effect: since so many experiments are run on the same data set, by chance, some will yield interesting (though meaningless) results

14

Model selection support in PyML

Is nested cross-validation difficult in PyML?

NO!

Let's take a look at PyML/classifiers/modelSelection.py

See also:
http://pyml.sourceforge.net/tutorial.html#model-selection

15