Laboratorio di Apprendimento Automatico

Fabio Aiolli
Università di Padova

Model Selection and Hold-out

- Most of the time, the learner is parametric. These parameters should be optimized by testing which values of the parameters yield the best effectiveness.

- Hold-out procedure
  1. A small subset of Tr, called the validation set (or hold-out set), denoted Va, is identified
  2. A classifier is learnt using examples in Tr-Va.
  3. Step 2 is performed with different values of the parameters, and tested against the hold-out sample

- In an operational setting, after parameter optimization, one typically re-trains the classifier on the entire training corpus, in order to boost effectiveness (debatable step!)

- It is possible to show that the evaluation performed in Step 2 gives an unbiased estimate of the error performed by a classifier learnt with the same parameters and with training set of cardinality $|Tr|-|Va|<|Tr|$

K-fold Cross Validation

- An alternative approach to model selection (and evaluation) is the K-fold cross-validation method
- K-fold CV procedure
  - K different classifiers $h_1, h_2, \ldots, h_k$ are built by partitioning the initial corpus Tr into k disjoint sets $Va_1, \ldots, Va_k$ and then iteratively applying the Hold-out approach on the k-pairs $\langle Tr_i = Tr - Va_i, Va_i \rangle$
  - Effectiveness is obtained by individually computing the effectiveness of $h_1, \ldots, h_k$, and then averaging the individual results
- The special case $k=|Tr|$ of k-fold cross-validation is called leave-one-out cross-validation

Evaluation for unbalanced data

- Classification accuracy:
  - usual in ML,
  - the proportion of correct decisions,
  - Not appropriate if the population rate of the class is low
- Precision, Recall and F_1
  - Better measures

Contingency Table

	Relevant	Not Relevant
Retrieved	True positives (**tp**)	False positives (**fp**)
Not Retrieved	False negatives (**fn**)	True negatives (**tn**)

$$\pi = \frac{tp}{tp+fp} \qquad \rho = \frac{tp}{tp+fn}$$

Why NOT using the accuracy $\alpha = \dfrac{tp+tn}{tp+fp+tn+fn}$?

Effectiveness for Binary Retrieval: Precision and Recall

If relevance is assumed to be binary-valued, effectiveness is typically measured as a combination of

- Precision: the "degree of soundness" of the system

$$\pi = Pr(Rel|Ret) = \frac{|\hat{Rel} \cap \hat{Ret}|}{|\hat{Ret}|}$$

- Recall: the "degree of completeness" of the system

$$\rho = Pr(Ret|Rel) = \frac{|\hat{Rel} \cap \hat{Ret}|}{|\hat{Rel}|}$$

F measure

- A measure that trades-off precision versus recall?
  F-measure (weighted harmonic mean of the precision and recall)

$$F = \frac{(\beta^2+1)\pi\rho}{\beta^2\pi+\rho}$$

$$F_{\beta=1} = \frac{2\pi\rho}{\pi+\rho}$$

$\beta < 1$ emphasizes precision!

ROC

- Receiver Operating Characteristic
- False Negative Rate plotted against the False Positive Rate

AUC (area under the ROC)

Efficient computation of AUC

- Assume *h(**x) returns a real quantity (larger values =>*** class 1)
- Sort **xi according to *h(xi). Number the sorted points*** from 1 to N such that r(i) = the rank of data point **xi**

AUC = probability that a randomly chosen example from class 1 ranks above a randomly chosen example from class 0 = the Wilcoxon-Mann-Whitney statistic

Adaboost con WEKA

- Ionosphere
- Iris