Data Mining

Classification VI - Evaluating Classifier Performance (Part A)

Dr. Jason T.L. Wang, Professor

Department of Computer Science

New Jersey Institute of Technology

Terminology

True positive (TP): the number of positive examples correctly predicted by the classification model.

True negative (TN): the number of negative examples correctly predicted by the classification model.

False positive (FP): the number of negative examples wrongly predicted as positive by the classification model.

False negative (FN): the number of positive examples wrongly predicted as negative by the classification model.

P: the number of positive examples $P = TP + FN$

N: the number of negative examples $N = TN + FP$

Terminology

True positive rate (TPR) or sensitivity: the fraction of positive examples predicted correctly by the model:

TPR = TP/(TP + FN) = TP/P

True negative rate (TNR) or specificity: the fraction of negative examples predicted correctly by the model:

TNR = TN/(TN + FP) = TN/N

False positive rate (FPR): fraction of negative examples predicted as positive:

FPR = FP/(TN + FP) = FP/N

False negative rate (FNR): fraction of positive examples predicted as negative:

FNR = FN/(TP + FN) = FN/P

Terminology

Recall (r) or sensitivity: $r = TP/(TP + FN) = TP/P$

Precision (p): $p = TP/(TP + FP)$

F_1 measure (F_1): $F_1 = (2 \times TP)/(2 \times TP + FP + FN)$

$= 2 \times (p \times r)/(p + r)$

Accuracy (Acc): $Acc = (TP + TN)/(TP + FP + FN + TN)$

$= (TP + TN)/(P + N)$

Error rate (Err): $Err = (FP + FN)/(TP + FP + FN + TN)$

$= (FP + FN)/(P + N)$

Evaluating Classifier Performance

Two methods:

10-fold cross validation – suitable for classifiers that predict labels (positive or negative); e.g. ID3, C4.5.

ROC/AUC – suitable for classifiers that produce continuously-valued outputs; e.g. support vector machines or neural networks.

10-Fold Cross Validation

Get a set **S** of examples where every example has a label (positive or negative).

Divide the set **S** into 10 folds/subsets of equal size where the union of the 10 folds is **S** and the intersection of fold i and fold j ($i \neq j$) is empty.

At run i, use fold i as the test dataset and the other nine folds together as the training dataset. Calculate the accuracy.

Take the average of the accuracies of all ten runs. This average is treated as the accuracy of the evaluated classifier.

End of

Evaluating Classifier Performance Module (Part A)