Semestralklausur zur Vorlesung

Maschinelles Lernen: Symbolische Ansätze WS 14/15

Prof. J. Fürnkranz
Technische Universität Darmstadt — Wintersemester 2014/2015
Termin: 17. 2. 2015

English translation

This is **not the official exam**, this is only meant to be an aid for students not speaking German. It is **not guaranteed** that the translation is complete or correct. In particular, this translation contains **no figures or tables**! Always **refer to the official exam** and use this only as a help.
Do not write your solutions on these pages! You may answer in English on the regular exam.

Aufgabe 1 Miscellaneous

1–a What is the purpose of the use of *kd*-trees with a *k*-NN classifier?

1–b Explain in words (without formulas) the concepts *false positives* and *false positive rate*.

1–c In which case can you maintain that you do not need any more additional training examples for the candidate elimination algorithm?

1–d What have all classifiers that are contained in the version space in common with respect to the training data? (exactly one answer is correct)

- They are all neither complete nor consistent
- They are complete but not necessarily consistent
- They are consistent but not necessarily complete
- They are all both complete and consistent
- The classifiers in the G-set are complete, the classifiers in the S-Set are consistent, for all others you can't say with certainty.

1–e You use bagging for training an ensemble of classifiers. For what purpose could you use *stacking* in this context?

1–f Both *boosting* and *bagging* can be realized by sampling from the original training data of size n a new training set of size n. What is the difference between the used sampling strategies?

1–g Assume a heuristic $h(x) \rightarrow [0, 1]$ which assigns a value between 0 and 1 to each rule, and a heuristic $g(x) = h^2(x)$, which squares the heuristic $h(.)$. Have h and g the same isometrics? Why (not)?

Aufgabe 2 Decision Trees

Given are the following 10 training instances for the prediction of T using the three nominal attributes L, M, and N.

2–a Given is the following measure for measuring the impurity of a set E with e_+ positive and e_- negative examples.

Define on the basis of this measure a gain heuristic analogously to the definition of information gain where entropy is used as a measure of impurity.

2–b Use the gain heuristic defined in a) for learning a decision tree from these data.

Remark: If you could not solve a), you can also use a measure that we discussed in the course (e.g., Gini index).

2–c Convert the resulting tree into a rule set and specify it. Which conditions of the rule set can be pruned without increasing the error on the training data?

2–d Which method would you use on this training set for evaluating your decision tree learner (exactly one answer is correct)

- 10-fold cross validation
- leave-one-out cross validation
- does not matter

Justify your answer.

Aufgabe 3 Multi-Class and Meta classification

Given is the following multi-class dataset:

3–a You want to tackle this dataset with a concept learner and use *pairwise classification* for this purpose. Give the training sets that you have to generate for the concept learner.

3–b As an alternative, error-correcting output codes are proposed to you, where you are expected to use the following coding matrix, in which the classes are coded in the order (a, b, c):

Answer the following questions:

1. How many binary classifiers do you have to train?
2. How many training examples for these binary classifiers are *in total* (across all datasets) labeled as positive, how many as negative?
3. Assume that the classifiers return the following class vector:
   Which of the classes (a, b, c) would you predict? Why?

3–c How would the coding matrix look like if you want to realize one-against-all classification with the help of ECOCs?

3–d You are using the examples $\{\#3, \#4, \#5\}$ in a first iteration of windowing. There, the following *decision list* is learned:

Which examples will be used as training set in the next iteration of windowing, and why?

Aufgabe 4 Evaluation

For the following 10 examples x is given their true class $h(x)$ as well as the probability $P(+|x)$ with which a learner will classify this example as positive.

4–a Draw the ROC curve for the above 10 examples.

Hint: The above checkered part of the paper is meant to be a drawing aid. You are not required to use it, and you can use it in any scale you wish. Don't forget to also draw the limits of the ROC space.

4–b Determine the area under the ROC curve (AUC).

4–c The area under ROC curve may be considered as a probabililty. Explain in words what this probability measures.

4–d You want to convert this ranker into a classifier K_t by introducing a threshold t for the probabilities $P(+|x)$ so that all examples with $P(+|x) > t$ are classified as positive, all others as negative. Determine from the ROC curve a possible threshold that maximizes the *accuracy* of the classifier, and compute this maximal accuracy.

4–e Compare the classifier $K_{0.5}$, which classifies all examples with $P(+|x) > 0.5$ as positive, to the classifier $K_{0.2}$, which classifies all examples with $P(+|x) > 0.2$ as positive. Starting with which cost ratio $\frac{c_-}{c_+}$ is it better to use $K_{0.2}$ than $K_{0.5}$?

Aufgabe 5 Association Rules

A fruit store makes the following observations about the buying habits of its customers:

Here is a list of all pairs of products that have been bought by at least 40% of the customers:

Hint: You can abbreviate the different fruits with their starting letters.

5–a Which candidates for itemsets of length 3 will be generated by APRIORI using the method discussed in the lecture?

Hint: The items in a shopping basket are sorted alphabetically.

5–b Which itemsets generated in a) can be removed from C_3 *before* their frequency has to be counted?

5–c Which itemsets form eventually the set S_3 of frequent itemsets of length 3? Specify the support of each member of S_3.

5–d Compute the lift of the association rule *Apfel* $\rightarrow$ *Erdbeere*.

5–e Which of the six elements of S_2 belongs to the positive border?

5–f Give an example of a 2-element itemset of the negative border.

5–g Give all association rules with $c_{min} = 0.8$ that can be formed from the itemset {*Apfel, Erdbeere, Clementine*}.