Machine Learning & Data Mining

Prova intermedia 26 aprile 2017

1) Concept learning

1.a) Describe in general what is the difference between biased and unbiased learning. Why should we prefer biased learners?

1.b) Which is the final version space computed by the Candidate Elimination Algorithm for learning the "EnjoySport" concept using the following training instances?

Sky	Temp	Humid	Wind	Water	Forecst	EnjoySpt
Sunny	Warm	Normal	Strong	Warm	Same	Yes
Sunny	Warm	High	Strong	Warm	Same	Yes
Rainy	Cold	High	Strong	Warm	Change	No
Sunny	Warm	High	Strong	Cool	Change	Yes

2) Decision trees

2.a) Describe the method of splitting based on Information Gain.

2.b) Which are the functions contained in the hypothesis space of the algorithms for learning decision trees?

2.c) Briefly describe the probabilistic method to handle missing values in the training instances used for learning a decision tree. Suppose we have the following 9 complete instances and 1 incomplete instance:

Refund	MarritalStatus	TaxableIncome	Class
yes	single	100k	no
no	married	200k	no
no	single	150	no
yes	married	110k	no
no	divorced	90k	yes
no	married	67k	no
yes	divorced	110k	no
no	single	89k	yes
no	married	76k	no
yes	???	59k	yes

Which is the result of splitting using attribute "*MarritalStatus*"?

3) Valutazione di algoritmi

What is the meaning of "Confidence Interval" when we use the sample error of a hypotheisis to estimate the true error of the hypothesis?

4) Reti Neurali

4.1) Specify the steps that are executed by the backpropagation algorithm for each epoch of the algorithm.

4.2) Indicate which are the typical termination conditions of the backpropagation algorithm.

5) Learning Bayesiano

5.1) From a practical point of view, why the Naive Bayes Classifier is more usable than the Optimal Bayes Classifier?

5.2) Which are the probabilities involved in the definition of Naive Bayes Classifier and how can they be estimated?