Machine Learning & Data Mining

Prova intermedia 10 maggio 2019

1) Concept learning

Describe how the Version Space generated as the output of the Candidate Elimination Algorithm is used to decide the classification value (Yes or NO) of a new instance X.

2) Alberi di decisione

2.a) Specify the main steps of the **(Hunt's) Algorithm for the construction of a decision tree.**
2.b) Describe how is defined the best split in terms of "best impurity reduction" and specify two alternative measures of impurity.
3.c) Specify the probabilistic classification for the new instance in the following scenario (explain the answer):

COME SI CLASSIFICA LA NUOVA ISTANZA?

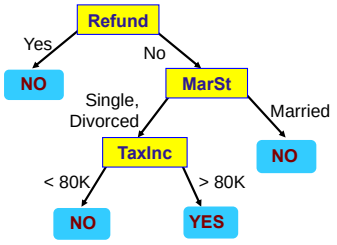
Tid	Refund	Marital Status	Taxable Income	Class
1	Yes	Single	125K	No
2	No	Married	100K	No
3	Yes	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	Yes	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	Yes	Married	75K	No

DECIMO RECORD INCOMPLETO

Tid	Refund	Marital Status	Taxable Income	Class
10	?	Single	90K	Yes

New record/instance:

Tid	Refund	Marital Status	Taxable Income	Class
11	No	?	85K	?



72

3) Valutazione di algoritmi

3.1) When we use the K-fold validation algorithm to compare two learning algorithms what are we estimating?
3.2) Give the definitions of Error rate, Accuracy, Recall.

4) Reti Neurali

Indicate the steps of the backpropation algorithm for a network with one hidden layer. If there are two hidden layes instead of one, what is the difference in the algorithm?

5) Learning Bayesiano

5.1) Suppose that we have a new instance x with possible classifications "yes" and "no", and three possible hypotheses such that:
P(h1|D) = 0.1, P(h2|D) = 0.5, P(h3|D) = 0.4
P("no"|h1) = 0, P("yes"|h1) = 1
P("no"|h2) = 0, P("yes"|h2) = 1
P("no"|h3) = 1, P("yes"|h3) = 0

Which is the value v* computed by the Optimal Bayes Classifier for this example? Why?

5.2) Which is the formula defining the Naïve Bayes Classifier? How can be estimated the probabilities in the formula?