Machine Learning & Data Mining

Prova scritta 13 Giugno 2017

1) **Concept learning**

Describe how new instances are classified using the version space generated by the Candidate-Eliminate algorithm. Does the Find-S algorithm always give the same classification as the Candidate-Eliminate algorithm?

2) **Decision trees**

2.a) Indicate when a classification problem is appropriate for using Decision Trees.

2.b) Splitting based on info gain tends to prefer attrinues with many values. Why can this be a problem? and how can we modify the info gain to avoid it?

2.c) Describe the method "Reduced-Error Pruning" to avoid overfitting in a decision tree.

3) **Evaluation of learning algorithms**

3.1) Using the concepts of error and accuracy, define when a hypothesis h overfits the traing data.

3.2) Define the "Recall" measure.

3.3) What is the "Bootstrap" method?

4) **Artificial neural networks**

4.1) What do we mean by "local error minimum" in the backpropagation algorithm?

4.2) What is the weight "momentum" in the backpropagation algorithm?

5) **Bayesian learning**

5.1) Specify which are the conditional probabilities that we need to estimate using the training data to build a Naive Bayes Classifier.

5.2) What is the problem of a standard Naive Bayes Classifier when some attribute value is missing in the training data?

6) **Clustering**

You are given two sets of 100 points that fall within the unit square. One set of points is arranged so that the points are uniformly spaced. The other set of points is generated from a uniform distribution over the unit square.

a) Is there a difference between the two sets of points?
b) If so, using k-means, which set of points will typically have a smaller SSE for K=10 clusters?
c) What will be the behavior of DBSCAN on these two data sets?

7) **Regression**

We are given the following 11 observations on variables x and y:

x	0	2	6	11	1	4	14	15	5	10	8
y	14	13.4	11.2	10.7	7.7	11.8	8.8	5.5	8.5	10	7.6

We want to see if a linear model gives a reasonable fit of the data, so we estimate the model with least squares. Compute the least squares estimates of w_0 and w_1.

Compute the proportion of the variance in y that is explained by the model (R2= 1-SSE/SST). Where SSE= Σe^2 and SST=$\Sigma(y-\mu(y))^2$

8) **Ensemble methods**

Explain the bias vs. variance tradeoff in statistical learning.
Explain the training and testing phases of the Random Forests algorithm and compare it with Boosting.

9) **SVM**

What is a kernel function? What is the kernel trick? Make an example of a commonly used kernel.

Consider the following linearly separable two-dimensional input vectors.

X	9	20	22	20	18	1	7	24	25	5	16	12
y	9	12	11	18	24	2	10	27	12	14	6	20
Class	A	B	B	B	B	A	A	B	B	A	A	B

Find the linear SVM that optimally separates the classes by maximizing the margin. Show the support vectors and compute the margin.