ESTIMATION OF DEPENDENCIES BASED ON SMALL NUMBER OF OBSERVATIONS.

Vladimir Vapnik

1. The main conceptual problem with estimating the dependencies from empirical data arises when the number of given observations is small. (The number of observations, ℓ, used in estimating the function from the set with VC-dimension h, is small if the ratio ℓ/h is small, say $\ell/h < 30$).

2. To do well in this situation, one has to use an induction principle that takes into account, along with the performance on the training set, the VC-dimension of the set of functions from which the decision function is chosen.

An important point is that, the VC-dimension of the set of functions, in general, does not coincide with the number of free parameters, or with the dimensionality of the input space. It is therefore possible to construct methods that generalize well even in a very high dimensional input space using a small number of observations.

3. To do the best, given a small sample size one also must try to solve only the problem which one really needs to solve, rather than some more general problem. Often, however, this is not easy.

For many applications (including financing) it is important to estimate the values of the function at the given points of interest, rather than to estimate the function itself. However, classical techniques first estimate the function (i.e. estimate the values of the function at an infinite number of points, constituting the area of definition of the function) and only then use the estimated function to find the values of the function at the given points of interest.

Therefore the classical technique solve a particular problem (estimating the values of a function at the points of interest) by first solving, as an intermediate problem, a more general one (estimating the function itself).

For small sample size there exist methods for estimating the values of a function at the points of interest that are not based on estimating the function itself. According to this theory, estimating the function at the given points of interest, and estimating the function itself are different problems. In the case of a small number of given observations, one must choose which problem needs to be solved.

4. The decision making problem needs even less information for its solution than the problem of estimating the values of a function at given points. Therefore, in situations with a small amount of information a direct decision making can be more precise than decision making based on estimating the results for alternatives. It is possible to construct algorithms for direct decision making.

5. Lastly I would like to describe the problem of learning to make an actions, whose solution is not based on the model estimating technique. The problem of learning to make an actions is a generalization of the decision making problem. It appears applicable to the many financing problems.

0-7803-2145-6/95 $4.00