T-61.5040 Oppivat mallit ja menetelmät
T-61.5040 Learning Models and Methods
Pajunen, Viitaniemi

Exercises 5, 16.2.2007

Problem 1. Consider the growth function $G(n) = \log \max_{Z_n} N(Z_n)$, that is, the log of maximum number of dichotomies for any set Z_n having n observations, using a given set of functions H. The growth function $G(n)$ is either

$$G(n) = n \log 2$$

or

$$G(n) \leq v(\log \frac{n}{v} + 1)$$

The integer v is defined by

$$\begin{aligned} G(v) &= v \log 2 \\ G(v+1) &< (v+1) \log 2. \end{aligned}$$

i) Why is the growth function linear at all points $n = 1, 2, \ldots, v$?

ii) If v is finite, what is $\lim \frac{G(n)}{n}$ when $n \to \infty$?

Problem 2.

Statistical Learning Theory gives a bound on the difference of the training error s and the average error c, conditional on c and the number of training points m. Assume you are predicting the last bit of a binary vector from a bag containing 2^{n-1} vectors with different $n-1$ first bits. Assume that your set of solutions H contains just one function h.

i) Given a training set of m vectors, what can we say about the average error c made by the function h when a vector is picked randomly from the bag including the training vectors?

ii) What can we say about $|c - s|$ on average over all training sets of size m? (demo)

Problem 3.

The bounds concerning average error c in Statistical Learning Theory are obtained from the likelihood $p(s|c, \ldots)$. If the *prior* on c would be constant, then the likelihood as a function of c would be proportional to the posterior $p(c|s, \ldots)$. However, SLT claims that one does not have to assume anything about the learning problem. Consider predicting bits $y_1, \ldots, y_n$ with vectors $x_1, \ldots, x_n$ as inputs with a function h so that $\hat{y}_i = h(x_i)$. Then the average error is $c = \frac{1}{n} \sum_i |y_i - h(x_i)|$.

i) What is a reasonable prior $p(c)$ if you have selected a solution h without assuming anything about the learning problem?

ii) Define a learning problem so that $y_1, \ldots, y_{10}$ are obtained from the results of spinning a roulette wheel at a casino. Assume that the result is a number that can only be black or red. Then $y_i \in \{\text{black}, \text{red}\}$. You have selected a function h which tells you which colour to play. Relying on SLT, you have decided that the average error c is uniformly distributed. The casino manager offers you a deal: for a cost of $2500 EUR$ you can play ten times, and if you win at least seven bets you will win 10000 EUR. Keeping your constant prior in mind, should you take this deal? Is the deal sensible for the casino manager, who uses the "correct"prior?

Problem 4. (demo)

In practice, it would be useful to know something about the average error c given a fixed training set. Consider binary classification and a set H containing only one h. Compare the probabilities for c *conditional to a fixed training set of size* m, for the values $c = 0.1$ and $c = 0.5$. Assume $s = 0.1$, $m = 100$, and $n = 1000$ (the number of input points). You will need to make assumptions about c: assume that you have no information to help you choose the function h.