Intelligent Data Analysis II

2. Exercise Sheet

Winter term 2015/2016 Handed out: 26.10.2015
Tutorial: 03.11.2015

Exercise 1 *Confidence Intervals*

We have evaluated a trained classifier f_θ on an independent test set, using zero-one loss as the evaluation criterion. It classified 76 of the 100 available test instances correctly, that is, the empirical risk is $\hat{R}_S = 0.24$.

1. What is the probability that the true risk of the classifier is lower than 0.25?
2. Compute a 95% two-sided confidence interval of the form $[\hat{R}_S - \epsilon, \hat{R}_S + \epsilon]$ for the risk estimate.

Exercise 2 *Sign Test, t-Test*

We have observed the following accuracy rates of two algorithms on ten data sets:

Data set	1	2	3	4	5	6	7	8	9	10
Algorithm 1	0.9	0.8	0.9	0.7	0.75	0.8	0.7	0.9	0.8	0.9
Algorithm 2	0.7	0.65	0.7	0.75	0.76	0.5	0.4	0.95	0.6	0.8

We would like to decide whether Algorithm 1 is preferable to Algorithm 2 or whether the null hypothesis holds that the probability of Algorithm 1 outperforming Algorithm 2 on a randomly chosen data set is 0.5.

1. What is the p-value of a sign test carried out on these results? Let's assume we evaluate the algorithms on k additional data sets and Algorithm 1 always outperforms Algorithm 2. How large would k have to be in order to reject the null hypothesis at $\alpha = 0.05$?
2. Would a paired t-test reject the null hypothesis at $\alpha = 0.05$? Compute the p-value of the paired t-test on these results.

We observe $n = 6$ measurements

$$X_1 = 0.05, X_2 = 0.6, X_3 = 0.45, X_4 = 0.35, X_5 = 0.21, X_6 = 0.55.$$

We have prior knowledge that the measurements are independently sampled from a uniform distribution, $X_i \sim U(0, \theta)$ for $1 \leq i \leq n$, and that the upper limit of the distribution is $\theta \leq 1$. However, we are unsure about the exact value of the upper limit θ.
We formulate a statistical test with the null hypothesis $H_0 : \theta = 1$. The test statistic is

$$T = \max_{i=1..n} X_i.$$

We will reject H_0 if $T < c$, for a threshold c that remains to be chosen.

1. We first study the cumulative distribution function of the test statistics, that is, we determine $p(T < c)$. Show that $p(T < c) = \left(\frac{c}{\theta}\right)^n$.

2. We would like to limit the Type I error rate to $\alpha = 0.05$. How do we have to choose c in order to achieve this?

3. Given the observed measurements, would the test reject the null hypothesis that $\theta = 1$ at $\alpha = 0.05$? What is the p-value?