Übung zur Vorlesung

Maschinelles Lernen 2

Sommersemester 2013

Abteilung Maschinelles Lernen
Institut für Softwaretechnik und theoretische Informatik
Fakultät IV, Technische Universität Berlin
Prof. Dr. Klaus-Robert Müller
Email: klaus-robert.mueller@tu-berlin.de

Exercise Sheet 7

Due **June 3rd**, 9 a.m. local time on ISIS

Relevant Dimensionality Estimation (RDE) was discussed in the lecture as a method to estimate the effective dimensionality of the data in the kernel feature space, and to get information about the noise level of the data. This week's task comprises the implementation of RDE and the analysis of some data sets.

Please work on the following tasks:

1. **(10 points)** Write a funktion `sincdata` with input arguments n and r, which draws and returns n data points from the following distributions:

$$X_i \sim \text{uniformly distributed between -4 and 4}$$
$$Y_i \sim \sin(\pi X_i)/(\pi X_i) + r\epsilon_i,$$

   where ϵ_i (representing noise) are normally distributed.

2. **(10 points)** Write a function `sinedata` with input arguments n and k, which draws and returns n data points from the following distribution:

$$X_i \sim \text{uniformly distributed between } -\pi \text{ and } \pi$$
$$Y_i \sim \sin(kX_i) + 0,3\,\epsilon_i.$$

   (The ϵ_i again are normally distributed.)

3. **(50 points)** Implement function `rde` based on the following log-likelihood estimator:

$$\hat{d} = \underset{1\le d\le \lfloor \frac{n}{2} \rfloor}{\arg\min} \frac{d}{n} \log \frac{1}{d} \sum_{i=1}^{d} s_i^2 + \frac{n-d}{n} \log \frac{1}{n-d} \sum_{i=d+1}^{n} s_i^2,$$

   where $s = (u_1^T Y, \dots u_n^T Y)$, $u_1, \dots, u_n$ represent the eigenvectors of the kernel matrix (sorted in order of the descending eigenvalues).

   The requested output is the estimated effective dimensionality $\hat{d}$ and $\hat{Y}$, which contains the projection of the original labels Y:

$$\hat{Y} = \sum_{i=1}^{\hat{d}} u_i u_i^T Y.$$

4. **(30 points)** Write a function `plot_fit`, which

   - performs an RDE analysis of sinc- and sine data for different parameters (e.g. noise levels r),
   - visualizes the data with labels, X, Y, and the projected (cleaned) labels $\hat{Y}$,
   - visualizes the estimated effective dimensionality.

(For your convenience, function stubs are provided via ISIS.)

Please ask questions in the ISIS discussion forums for Machine Learning 2:
`https://www.isis.tu-berlin.de/course/view.php?id=8005`