

Maschinelles Lernen 2

Sommersemester 2008

Blatt 9

Abgabe 17. Juni 2008 in der Vorlesung. Praktische Übungsaufgaben über PASS abgeben (<https://ml01.zrz.tu-berlin.de/~mikio/pass.pl?conf=blatt9.conf>). Verwende matlab (/home/ml/ml/bin), oder octave (<http://www.octave.org> frei verfügbar).

Aufgaben

In der Vorlesung wurde die RDE-Methode besprochen, um die effektive Dimensionalität im Kernmerkmalsraum zu bestimmen. Auf diesem Aufgabenblatt soll die Methode implementiert und einige Datensätze untersucht werden.

1. (10 Punkte) Implementiere RDE basierend auf der Leave-One-Out Kreuzvalidierung. Zur Erinnerung, der Fehler ist

$$\frac{1}{n} \sum_{i=1}^n \left(\frac{Y_i - \hat{Y}_i}{1 - S_{ii}} \right)^2,$$

wobei $\hat{Y} = SY$, $S = \sum_{i=1}^d u_i u_i^T$ und u_i die Eigenvektoren der Kernmatrix in absteigender Sortierung des dazugehörigen Eigenwerts ist.

Schreibe eine Funktion `rde.m` mit Argumenten Kernmatrix K , Labelvektor Y , die den obigen Fehler für alle $d = 1, \dots, n$ zurückgibt, wobei n die Anzahl der Datenpunkte ist. Schreibe die Methode so, dass für jedes d der Aufwand nur linear in der Anzahl der Beispiele ist.

2. (15 Punkte) Schreibe eine Funktion `sincdata.m` mit Argumenten n und s , die n Punkte von folgenden Verteilungen generiert:

$$\begin{aligned} X_i &\sim \text{gleichverteilt von } -4, 4 \\ Y_i &\sim \sin(\pi X_i) / (\pi X_i) + s \varepsilon_i, \end{aligned}$$

wobei die ε_i normalverteilt sind.

Schreibe eine Funktion `aufgabe2.m`, die `sincdata` Datensätze mit $n = 100$ und $s = 0.01, 0.1, 0.5$ erzeugt und

- Für einen Gaußschen Kern

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2w}\right)$$

für die Kernweiten $w = 0.01, 0.5$ und 10 ,

- jeweils $|u_i^T Y|$ für $i = 1, \dots, n$,
 - den mit `rde` berechneten Fehler für $d = 1, \dots, n$,
 - sowie \hat{Y}_i gegen die X_i plottet.
3. (15 Punkte) Schreibe eine Funktion `sinedata.m` mit Argumenten n und k , die n Punkte von der Verteilung

$$\begin{aligned} X_i &\sim \text{gleichverteilt von } -\pi, \pi \\ Y_i &\sim \sin(kX_i) + 0,3\varepsilon_i. \end{aligned}$$

zieht, wobei die ε_i normalverteilt sind.

Schreibe eine Funktion `aufgabe3.m`, die mit `sinedata` Datensätze mit $n = 100$ und $k = 1, 10, 100$ erzeugt und wie in Aufgabe 2 die Skalarprodukte, die `rde` Fehler und die entrauschen Labels plottet.