

**Machine Learning and Data Mining**  
Summer 2015  
**Exercise Sheet 9**

*Presentation of Solutions to the Exercise Sheet on the 24.06.2015*

**Exercise 9-1** Modellvergleich

Vergleichen Sie Modelle der Regression mit Basisfunktionen. Die Vorhersage für einen Datenpunkt  $\mathbf{x}_i \in \mathbb{R}$  sei gegeben durch:

$$f(\mathbf{x}_i, \mathbf{w}) = \sum_{j=1}^{M_\Phi} w_j \phi_j(\mathbf{x}_i)$$

Verwenden Sie die PLS-Lösung  $\hat{\mathbf{w}} = (\Phi^T \Phi + \lambda I)^{-1} \Phi^T \mathbf{y}$  mit  $\Phi_{i,j} = \phi_j(\mathbf{x}_i) = \mathbf{x}_i^{j-1}$ . Gegeben sei der Datensatz  $\mathbf{X}, \mathbf{y}$  der Größe  $N = 10$ , basierend auf einer Rauschvarianz von  $\sigma^2 = 0.25$ :

$\mathbf{X}$	0.3	0.4	0.8	1.5	1.8	3.6	4	4.3	4.6	5
$\mathbf{y}$	7	4.7	0.6	-1.1	-0.3	4.6	5.5	5.7	3.1	-0.3

Es soll das beste Modell für Basisfunktionen mit  $M_\Phi \in \{1, \dots, 6\}$  bestimmt werden. Als Loss-Funktion nehmen Sie im folgenden den mittleren quadratischen Fehler (MSE).

- Bestimmen Sie das beste Modell durch Kreuzvalidierung (5-fach und 10-fach). Unterstützen die paarweisen Tests aus der Vorlesung die Entscheidung des MSE? Welchen Einfluss hat der  $\lambda$ -Parameter?
- Kommen Sie mit den frequentistischen Verfahren ( $C_p$  Statistik und AIC) und dem Bayes'schen Verfahren (BIC) zu den gleichen Schlüssen?
- Welchen Einfluss hat die Datengröße  $N$ , wenn Sie einen vergleichbaren Datensatz für  $N = \{100, 1000\}$  simulieren?