Maschinelles Lernen

6. Übung

Prof. Tobias Scheffer
Dr. Niels Landwehr
Christoph Sawade
Paul Prasse

WS09/10

Ausgabe am: 01.12.09
Besprechung am: 08.12.09

Aufgabe 1 (1/3 Punkt):
Nehmen wir an, wir haben eine Menge von Beispielen $x_1, \ldots, x_n$, unabhängig aus einer $N(\mu, 1)$ normalverteilten Grundgesamtheit gezogen. Die Dichte einer univariaten Normalverteilung ist:

$$p(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Wir möchten eine Parameterschätzung für den Parameter μ durchführen. Leiten sie den Maximum-Likelihood-Schätzer für den Parameter μ her. Die Likelihood der Daten ist $P(x_1, \ldots, x_n|\mu, \sigma^2) = \prod_{i=1}^{n} p(x_i|\mu, \sigma^2)$. Hinweis: Da log() eine streng monotone Funktion ist, befindet sich an einem Maximum von $\log P(x_1, \ldots, x_n|\mu, \sigma^2)$ auch ein Maximum von $P(x_1, \ldots, x_n|\mu, \sigma^2)$.

Aufgabe 2 (1/3 Punkt):
Wir betrachten die Bayessche Regression für 2-dimensionale Eingabevektoren $x_1 = [2, 1]^\top$, $x_2 = [-4, 1]^\top$, $x_3 = [0, 1]^\top$ und $x_4 = [4, 1]^\top$, wobei wir annehmen dass die zweite Komponente der Eingaben immer konstant 1 ist. Zusätzlich sind folgende Werte für die Zielvariable gegeben: $y_1 = 1$, $y_2 = -1$, $y_3 = 0$ und $y_4 = 2$. Bestimmen sie $\bar{w}_1, \bar{w}_2$ bzw. w, c; einmal unter Verwendung der Formeln für die allgmeinen Bayesschen Regression (Folien ab Seite 89 zu Bayessches Lernen) mit $\sigma^2 = 2$ und $\Sigma_p^{-1} = 0$, und einmal unter Verwendung der Gleichungen der 1-dimensionalen linearen Regression (Folien ab Seite 50 zu Entscheidungsbäumen). Welcher Zusammenhang besteht zwischen der MAP-Hypothese der Bayesschen Regression und der 1-dimensionalen linearen Regression? Welchen praktischen Nutzen könnte die Varianz der Bayes-optimalen Lösung haben?

Aufgabe 3 (1/3 Punkt):
Sie haben in der Vorlesung die Bayes'sche Regression kennengelernt. Nun sei angenommen, dass ihre Trainingslabels y_i sich um ein Gauß-verteiltes Rauschen von den echten Datenpunkten (x_i, t_i) unterscheiden:

$$y_i = t_i + \varepsilon_i, \text{ mit } \varepsilon_i \sim N(0, \sigma_i^2).$$

Leiten Sie den ML-Schätzer für den Parameter w her. Inwiefern hat das Rauschen Einfluss auf die optimale Lösung? Betrachten Sie einen Datensatz in dem jeder Datenpunkt mit einem instanzspezifischen Gewicht w_i versehen wurde. Geben Sie zwei unterschiedliche Interpretationen für diese Umgewichtung an.