Ludwig-Maximilians-Universität München 15.06.2011
Institut für Informatik
Dr. Volker Tresp
Franz Graf, Marisa Petri

Maschinelles Lernen und Data Mining

Sommersemester 2011

Übungsblatt 5

Besprechung des Übungsblattes am 24.06.2011 (der Termin am 23.06. entfällt)

Aufgabe 5-1 Generatives Modell

a) Wenn $P(c = j)$ und $P(\mathbf{x}|c = j)$ bekannt sind, lässt sich der optimale Klassifikator nach der Bayes'schen Regel berechnen. Geben Sie ihn als Entscheidungsfunktion an.

b) Nun sei $P(\mathbf{x}|c = j)$ für alle j mit identischer Kovarianz aber unterschiedlichen Zentren normalverteilt. Formulieren Sie das Problem aus.

Hinweis: Die mehrdimensionale Normalverteilung ist definiert als

$$\mathcal{N}(\mathbf{x}_i|c = j) = \frac{1}{(2\pi)^{\frac{p}{2}} \, |\Sigma|^{\frac{1}{2}}} e^{\left(-\frac{1}{2}(\mathbf{x}_i-\mu_j)^T\Sigma^{-1}(\mathbf{x}_i-\mu_j)\right)} ,$$

wobei μ_j der Mittelwertsvektor zu Klasse j, Σ die Kovarianzmatrix über alle Klassen, und $|\Sigma|$ die Determinante von Σ ist.

c) Dieses Problem lässt sich optimieren zu einem Schätzer auf den μ_j und Σ:

$$\hat{\mu}_j = \frac{1}{N_j} \sum_{i:y_i=j} \mathbf{x}_i$$

$$\hat{\Sigma} = \frac{1}{N - M} \sum_{j=1}^{C} \sum_{i:y_i=j} (\mathbf{x}_i - \hat{\mu}_j)(\mathbf{x}_i - \hat{\mu}_j)^T ,$$

wobei N die Größe des Trainingdatensatzes ist, M die Dimension der Daten und N_j die Anzahl der Trainingsdaten, die zu Klasse j gehören. Die Grund-Wahrscheinlichkeiten für die Klassen, $P(c = j)$, lassen sich anhand ihrer Häufigkeit im Datensatz abschätzen.

Trainieren Sie das in b) entwickelte Modell auf den Daten aus `bayesianData.txt` und bestimmen Sie anschließend den Trainingsfehler. (Die Datei besteht aus 2 Tab-separierten Variablenspalten mit vorangehender Klassenlabelspalte.)

b.w.

Aufgabe 5-2 Wdh. Vektor Calculus

Eine Funktion $g(\mathbf{x}) : \mathbb{R}^n \rightarrow \mathbb{R}$ über einen Vektor $\mathbf{x} \in \mathbb{R}^n$, die nach $\mathbf{x}$ abgeleitet wird ist:

$$\frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}_1} \\ \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}_2} \\ \vdots \\ \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}_n} \end{bmatrix} .$$

Berechnen Sie $\frac{\partial \langle \mathbf{x}, \mathbf{x} \rangle}{\partial \mathbf{x}}$, wobei $\langle \mathbf{x}, \mathbf{x} \rangle$ das Standard-Skalarprodukt von $\mathbf{x}$ mit sich selbst ist.

Aufgabe 5-3 Lagrange Multiplikatoren

Wir betrachten ein Optimierungsproblem auf $x \in \mathbb{R}^n$ der Form $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ und $g_j : \mathbb{R}^n \rightarrow \mathbb{R}$:

$$\begin{array}{lll} \min_{x \in \mathbb{R}^n} & f_0(x) & \\ \text{so dass} & f_i(x) \leq 0 & \text{für } i = 1, \ldots, m \\ & g_j(x) = 0 & \text{für } j = 1, \ldots, p \end{array}$$

f_0 ist die Zielfunktion, f_i und g_j sind Nebenbedingungen (für Ungleichungsbedingungen und exakte Bedingungen). Eine Lagrangefunktion nimmt diese Nebenbedingungen mit in die Zielfunktion auf und optimiert dadurch innerhalb der angegebenen Grenzen:

$$\mathcal{L}(x, \gamma, \lambda) = f_0(x) + \sum_{i=1}^{m} \gamma_i f_i(x) + \sum_{j=1}^{p} \lambda_j g_j(x) ,$$

wobei γ und λ reelle Vektoren, die *Lagrange Multiplikatoren* sind. Sie werden auch als duale Variablen bezeichnet. Alle γ_i müssen ≥ 0 sein, die λ_j sind frei wählbar. In einfacheren Problemen kann man sie jeweils durch Minimierung nach den Zielparametern (also x) eindeutig bestimmen.

Optimieren Sie die folgenden Probleme für $n = 2$ mithilfe eines Lagrange-Termes für Zahlen $x_1, x_2 \in \mathbb{R}$, deren Summe 20 ist,

a) wenn $x_1 \cdot x_2$ maximal sein soll.

b) wenn $x_1^2 + x_2^2$ minimal sein soll.

c) wenn $e^{-(5x_1 - x_2)^2}$ maximal sein soll.

Stellen Sie hierzu zunächst jeweils den passenden Lagrange-Term auf, und minimieren Sie diesen nach den Zielparametern x_1 und x_2. Danach darf auch wieder die Eingangsbedingung ausgenutzt werden.