Übung zur Vorlesung

Maschinelles Lernen 2

Sommersemester 2008

Abteilung Maschinelles Lernen
Institut für Softwaretechnik und theoretische Informatik
Fakultät IV, Technische Universität Berlin
Prof. Dr. Klaus-Robert Müller
http://ml.cs.tu-berlin.de

Blatt 1

Abgabe: 22. April 2007, bis 12.15 h bei Nicole Krämer

Aufgaben

1. **Kerndarstellung (12 Punkte)** Wir betrachten zwei Datensätze $x_1, \ldots, x_n \in \mathbb{R}^d$ und $z_1, \ldots, z_n \in \mathbb{R}^p$, und nehmen an, dass beide Datensätze zentriert sind, d.h.

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 0 \quad , \quad \overline{z} = \frac{1}{n}\sum_{i=1}^{n} z_i = 0 \,.$$

Die Datensätze werden durch die beiden Matrizen

$$X = (x_1, \ldots, x_n) \in \mathbb{R}^{d\times n} \quad , \quad Z = (z_1, \ldots, z_n) \in \mathbb{R}^{p\times n}$$

dargestellt. Ziel der Aufgabe ist es, für das Optimierungsproblem

$$\max_{w\in\mathbb{R}^d, q\in\mathbb{R}^p} \quad \mathrm{cov}(X^\top w, Z^\top q) \tag{1}$$

$$\text{Nebenbedingung} \quad \|w\|^2 = \|q\|^2 = 1 \tag{2}$$

eine Kerndarstellung zu finden und die Lösung mit Hilfe von Kernen auszudrücken.

(a) Zeige, dass die Vektoren w^* und q^*, die das Optimierungsproblem (1) und (2) lösen, eine Linearkombination der Daten sind, d.h.

$$\exists \alpha^*, \beta^* \in \mathbb{R}^n : \qquad w^* = \sum_{i=1}^{n} \alpha_i^* x_i$$

$$q^* = \sum_{i=1}^{n} \beta_i^* z_i$$

(b) Stelle das Optimierungsproblem (1) und (2) als ein Problem in den Variablen α und β dar und verwende nur die Kernmatrizen

$$K_X = X^\top X \in \mathbb{R}^{n\times n} \qquad K_Z = Z^\top Z \in \mathbb{R}^{n\times n}$$

(c) Bestimme die optimalen Lösungen α^* und β^*.

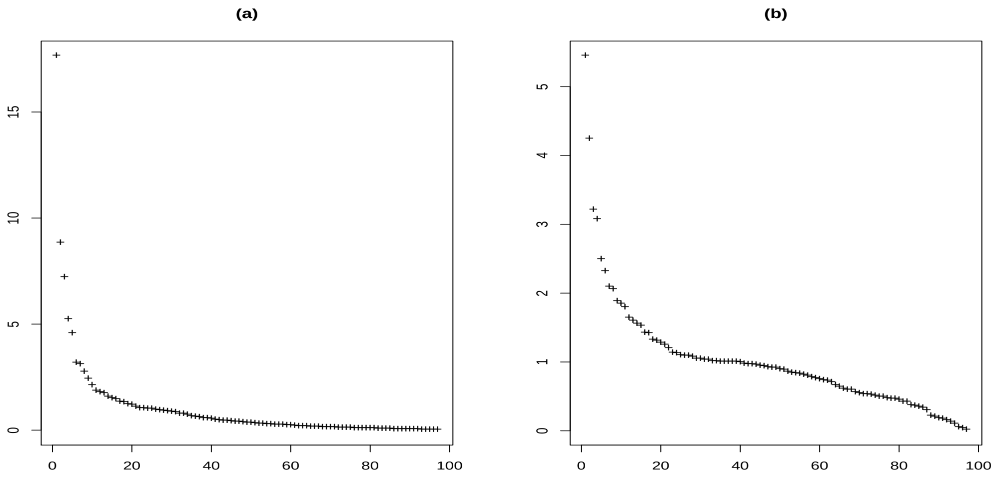
2. **Spektrum von Gaußkernen (8 Punkte)** Der Gaußkern mit Kernbreite $\sigma > 0$ ist definiert als

$$k_\sigma(x, x') \quad = \quad \exp\left(-\frac{\|x - x'\|^2}{\sigma^2}\right) \tag{3}$$

Für einen Datensatz mit n Beispielen $x_1, \ldots, x_n$ definiert die Kernfunktion eine $n \times n$ Matrix

$$K_\sigma \quad = \quad (k_\sigma(x_i, x_j))_{i,j=1,\ldots,n} \,.$$

In der folgenden Graphik sind für einen Datensatz der Größe 98 die Eigenwerte $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$ der Kernmatrix dargestellt. Auf der x-Achse is der Index i abgetragen, auf der y-Achse der Wert λ_i. Die beiden Abbildungen korrespondieren zu unterschiedlichen Kernbreiten $\sigma_1 > \sigma_2$.



(a) Welche der Kernbreiten $\sigma_1 > \sigma_2$ gehört zur linken Graphik? Begründe Deine Antwort.

(b) Welcher der beiden Kerne in der Abbildung overfittet stärker? Begründe Deine Antwort.

3. **Geometrie im Featureraum (8 Punkte)** Wir betrachten wieder den Gaußkern in (3), zu dem eine Featureabbildung $\Phi : \mathbb{R}^d \to \mathcal{F}$ assoziiert ist.

   (a) In welchem Bereich liegt die Länge eines Bild-Vektors im Featureraum?

   (b) In welchem Bereich liegt der Winkel zwischen zwei Bild-Vektoren im Featureraum?

   (c) In welchem Bereich liegt der Abstand zwischen zwei Bild-Vektoren im Featureraum?

   Gib in allen drei Fällen den kleinstmöglichen Bereich an und begründe Deine Entscheidung.

4. **Maximum-A-Posteriori (MAP) -Schätzer (12 Punkte)**

   (a) Zeige, dass der MAP-Schätzer im linearen Regressionsmodell äquivalent zum Ridge-Schätzer ist.

   (b) Zeige, dass der MAP-Schätzer für die logistische Regression die Lösung des Optimierungsproblems

   $$\operatorname{argmin}_w \qquad \left\{ \lambda \|w\|^2 + \sum_{i=1}^{n} \log\left(1 + \exp\left(-y_i \left(w^\top \Phi(x_i) + b\right)\right)\right) \right\}$$

   ist.

Hinweis:

- In Aufgabenteil 1 ist die empirische Kovarianz zwischen zwei Vektoren $u, v \in \mathbb{R}^n$ definiert als

$$\operatorname{cov}(u, v) = \frac{1}{n} \sum_{i=1}^{n} (u_i - \overline{u})(v_i - \overline{v}) = \frac{1}{n} \sum_{i=1}^{n} u_i v_i - \overline{u}\overline{v}\,.$$