

MUSTERERKENNUNG, 10. SEMINAR – KERNEL-PCA, HINGE-LOSS

Aufgabe 1. In der Vorlesung wurde die Aufgabe der Hauptkomponentenanalyse wie folgt definiert (wir bezeichnen das als „Variante A“). Gegeben sei eine Lernstichprobe der Muster $L = (x^1, x^2 \dots x^l)$, $x \in \mathbb{R}^n$. Gesucht wird der Richtungsvektor $e \in \mathbb{R}^n$, der die Summe der quadratischen Längen der Projektionen (die Streuung entlang e) maximiert:

$$\sum_i \langle x^i, e \rangle^2 \rightarrow \max_e$$

s.t. $\|e\| = 1$

Eine alternative Vorgehensweise ist („Variante B“), nach dem Richtungsvektor so zu suchen, dass die Summe der quadratischen Abstände zwischen Datenpunkten und ihren Projektionen auf e minimiert wird. Formulieren Sie die zu optimierende Zielfunktion. Zeigen Sie, dass diese zwei Varianten äquivalent sind.

Aufgabe 2. Vorher wurde immer der Fall betrachtet, in dem die Daten bereits zentriert sind, d.h. $\sum_i x^i = 0$. Wir modifizieren die Aufgabe, indem wir eine Verschiebung der Daten μ explizit als unbekannter Parameter betrachten. Zum Beispiel, lautet die in der Variante A zu lösende Aufgabe in diesem Fall

$$\sum_i \langle x^i - \mu, e \rangle^2 \rightarrow \max_{e, \mu}$$

s.t. $\|e\| = 1$.

a) Zeigen Sie, dass die Varianten A und B nicht mehr äquivalent sind. Welche davon ist „die richtige“?

b) Zeigen Sie, dass die Lösung bezüglich μ in der richtigen Variante nicht eindeutig ist, d.h. es existieren mehrere bezüglich μ optimale Lösungen. Zeigen Sie, dass der Mittelpunkt $\mu = 1/l \sum_i x^i$ eine davon ist.

Aufgabe 3. Man betrachte die Methode der Kernel-PCA, d.h. die Hauptkomponente wird im Merkmalsraum gesucht. Dabei sollen die Abbilder der Datenpunkte im Merkmalsraum zentriert werden, d.h. Abbilder $\Phi(x^i)$ werden durch $\Phi(x^i) - \mu$ ersetzt (mit $\mu = 1/l \sum_i \Phi(x_i)$, siehe vorige Aufgabe). Das Problem besteht darin, dass der Mittelpunkt im Merkmalsraum μ nicht explizit berechnet werden kann, denn die Abbildung Φ wird meist überhaupt nicht explizit definiert. Zeigen Sie, wie sich das Problem umgehen lässt, indem die entsprechende Aufgabe mithilfe nur Skalarprodukte definiert wird. Schlussfolgern Sie die Anwendbarkeit der Kernels für den Fall, dass die Daten im Merkmalsraum zentriert werden müssen.

Aufgabe 4. Zur Klassifikation der Muster $x \in \mathbb{R}$ soll die Schwellwertentscheidung verwendet werden, d.h. $y = \text{sgn}(x - \theta)$ mit ± 1 -Kodierung der Klassen und einem Schwellwert θ . Zum Anlernen des unbekannten Schwellwerts steht die folgende klassifizierte Lernstichprobe zur Verfügung: $L = ((x^l, y^l) \dots) = ((0, -1), (3, +1), (6, -1))$ – sie besteht aus drei Punkten und ist nicht separierbar.

a) Zeichnen Sie den Graph des Empirischen Risikos in Abhängigkeit von θ .

b) Um den optimalen Schwellwert zu finden, wird das Empirische Risiko durch Hinge-Loss ersetzt. Zeichnen Sie den Graph des Hinge-Loss in Abhängigkeit von θ . Vergleichen Sie die beiden Funktionen.