

MUSTERERKENNUNG, KERNEL-PCA

Man erinnere an die „gewöhnliche“ PCA. Gegeben sei Datenpunkte $x_j \in \mathbb{R}^n$, $j = 1 \dots l$. Man suche nach einem Richtungsvektor v so dass die Streuung der Projektionen der Datenpunkte auf diesen Vektor maximal ist. Diese Aufgabe wird wie folgt gelöst. Zunächst wird die Kovarianzmatrix berechnet:

$$Cov = \frac{1}{l} \sum_{j=1}^l x_j x_j^T \quad (1)$$

Danach wird der Eigenvektor v , d.h.

$$\lambda v = Cov \cdot v \quad (2)$$

zum größten Eigenwert λ gesucht (siehe Vorlesung „Bildverarbeitung: Bildmerkmale“, Folien 12-13).

Man kann zeigen, dass sich der optimale Vektor v als lineare Kombination der Datenpunkte ausdrücken lässt, d.h.

$$v = \sum_{i=1}^l x_i \alpha_i \quad (3)$$

mit $\alpha_i \in \mathbb{R}$. Mit anderen Worten, liegt der Vektor v in einem Unterraum, der von den Datenpunkten x_i aufgespannt ist. Für eine gewöhnliche PCA ist das oft irrelevant, denn es gibt (normalerweise) viel mehr Datenpunkte als die Dimension des Raumes, d.h. $l \gg n$. Das wird aber wichtig sein, wenn später die Dimension des Merkmalsraum (in dem PCA durchgeführt werden soll) sehr groß oder sogar unendlich sein kann.

Jetzt wird die Aufgabe zunächst etwas „verkompliziert“. Statt (2) zu lösen, wird

$$\lambda \cdot (x_k^T \cdot v) = x_k^T \cdot Cov \cdot v \quad \forall k \quad (4)$$

betrachtet, d.h. die Projektionen des optimalen Vektors v auf alle Datenpunkte x_k . Für jeden einzelnen Vektor x_k (an sich genommen) folgt die obige Formel aus (2) direkt. Weniger trivial ist die Aussage dass (2) und (4) zu einander äquivalent sind, d.h. v ist tatsächlich ein Eigenvektor von Cov , wenn sich alle Projektionen von v (nach der Multiplikation mit Cov) in einen und denselben Faktor λ ändern. Dies gilt nur, wenn (3)

erfüllt ist (hier gerade der Fall). Setzt man jetzt (1) und (3) in (4) ein, so erhält man

$$\begin{aligned}\lambda \cdot (x_k^T \cdot \sum_i x_i \alpha_i) &= x_k^T \cdot \frac{1}{l} \sum_j x_j x_j^T \cdot \sum_i x_i \alpha_i \quad \forall k, \\ \lambda l \sum_i \alpha_i x_k^T x_i &= \sum_i \alpha_i \sum_j x_k^T x_j x_j^T x_i \quad \forall k.\end{aligned}\tag{5}$$

Sei K die Matrix aller paarweisen Skalarprodukte, d.h. $K_{ij} = x_i^T x_j$. Somit lässt sich die obige Formel wie folgt schreiben:

$$\lambda l \sum_i \alpha_i K_{ki} = \sum_i \alpha_i \sum_j K_{kj} K_{ji},\tag{6}$$

und in der Matrixform (alle Gleichungen zusammen)

$$\lambda l K \alpha = K^2 \alpha\tag{7}$$

oder äquivalent

$$\lambda l \alpha = K \alpha\tag{8}$$

mit dem unbekannten Vektor $\alpha = (\alpha_1, \alpha_2 \dots \alpha_l)$ und der Matrix der Skalarprodukte K . Das ist wieder ein Eigenwert Problem genauso wie (2). Der Unterschied ist nur, dass statt Kovarianzmatrix Cov die Matrix der Skalarprodukte K genutzt wird.

Konsequenz: PCA lässt sich mithilfe nur Skalarprodukte ausdrücken!!!

Kernel-PCA: statt Skalarprodukte im ursprünglichen Raum werden Skalarprodukte im Merkmalsraum betrachtet, d.h.

$$K_{ij} = \langle \Phi(x_i), \Phi(x_j) \rangle.\tag{9}$$

Alles andere bleibt dasselbe. Wie bei den Kernels üblich ist, wird dabei die Abbildung Φ nicht explizit benötigt, wenn eine Kernelfunktion $\kappa(x, x')$ zur Verfügung steht, die den Skalarprodukt (9) realisiert.

P.S. Im Originalpaper (siehe Link auf der Seite) werden alle Schritte (1) – (8) gleich im Merkmalsraum betrachtet. Hier wurden alle Ableitungsschritte für „gewöhnliche“ PCA durchgeführt und nur zum Schluss das Ergebnis auf Merkmalsräume verallgemeinert.