Bildverarbeitung: Momente, Hauptkomponentenanalyse

D. Schlesinger – TUD/INF/KI/IS

Momente

Momente sind bestimmte Charakteristika der Wahrscheinlichkeitsverteilungen.

Sei $p(x)$ eine Wahrscheinlichkeitsverteilung einer reelwertigen Größe $x \in \mathbb{R}$.
Ein **Gewöhnlicher** Moment k-ter Ordnung um η ist

$$m_k(\eta) = \mathbb{E}\Big((x-\eta)^k\Big) = \sum_x p(x) \cdot (x-\eta)^k$$

Ein **Absoluter** Moment k-ter Ordnung um η ist

$$M_k(\eta) = \mathbb{E}\Big(|x-\eta|^k\Big) = \sum_x p(x) \cdot |x-\eta|^k$$

$\eta = 0$: Momente um 0 (bezeichnet mit m_k). **Mittelwert** ist m_1.

Zentrale Momente (bezeichnet als μ_k)– Momente um den Mittelwert ($m_1(0) = \mu_1$)
Varianz: zentraler Moment zweiter Ordnung

$$\mu_2 = \sum_x p(x) \cdot (x-\mu_1)^2$$

Interessant: $\mu_2 = m_2 - m_1^2$ (Beweis auf der Tafel).

Standardabweichung ist $\sigma = \sqrt{\mu_2}$

Spezielle Moment-basierte Maße

Schiefe:

$$v = \frac{m_3}{\sigma^3} = \frac{\sum_x p(x) \cdot (x - \mu_1)^3}{\left(\sum_x p(x) \cdot (x - \mu_1)^2\right)^{3/2}}$$

ist (u.a.) ein Maß für die Symmetrie der Wahrscheinlichkeitsverteilung zum Mittelwert. Es wird oft als „Abweichung“ von der Normalverteilung (bei der $v = 0$ ist) benutzt.

Interessant: Die Schiefe ist invariant unter linearer Transformation: $v(aX + b) = v(X)$

Wölbung:

$$\beta = \frac{m_4}{\sigma^4}$$

(nennt man auch **Kurtosis**) beschreibt die Peakhaftigkeit einer Verteilungsfunktion.

– Sehr oft ist eine WV vollständig durch alle ihre Momente definiert.

– Manche WV sind durch bestimmte Momente vollständig definiert.

  Beispiel: eine Gaussche WV ist durch Mittelwert und Varianz definiert – alle anderen Momente ergeben sich daraus.

– Es gibt WV, für die keine Momente existieren.

Momente vektorieller Größen

Gegeben sei die Wahrscheinlichkeitsverteilung $p(x)$ einer vektoriellen Größe $x \in \mathbb{R}^n$

Die **Kovarianz** für zwei Komponenten x_i und x_j ist

$$cov(x_i, x_j) = \sum_x p(x) \cdot (x_i - \nu)(x_j - \eta)$$

mit den jeweiligen Mittelwerten $\nu = \sum_x p(x) \cdot x_i$ und $\eta = \sum_x p(x) \cdot x_j$.

Kovarianzmatrix: $cov_{ij} = cov(x_i, x_j)$
Symmetrisch, positiv semidefinit, die Diagonalelemente sind die Varianzen.
Sind die Komponenten von x unabhängig, so ist die Matrix diagonal,
d.h. $cov(x_i, x_j) = 0$ für $i \neq j$ (Beweis auf der Tafel).
Achtung! Das umgekehrte gilt im Allgemeinen nicht.

In der Praxis werden oft die Momente einer Wahrscheinlichkeitsverteilung anhand einer Lernstichprobe $L = (x^1, x^2, \ldots, x^{|L|})$ ermittelt. Für die Kovarianzmatrix ergibt sich dabei

$$cov = \frac{1}{|L|} \sum_l (x^l - \mu) \otimes (x^l - \mu)$$

$\otimes$ ist Kreuzprodukt, μ ist der Mittelwert $\mu = \frac{1}{|L|} \sum_l x^l$.

Ein Anwendungsbeispiel

Die Idee – die Kovarianzmatrix beschreibt (in einem gewissen Sinne) die Form der Wahrscheinlichkeitsverteilung (Varianzen, Abhängigkeiten).
$\rightarrow$ dasselbe kann als quantitative Charakteristika eines Objektes verwendet werden.

Die „Lernstichprobe“ besteht aus Pixeln, die einem Objekt gehören.
Die Variablen sind die x und y Koordinaten.

- Die Mittelwerte entsprechen der Lage des Objektes.
- Die Varianzen entsprechen den charakteristischen Größen entlang der Achsen.
- Die Achsensymmetrische Objekte haben Null Kovarianzen
- Die Kovarianzen entsprechen der Rotation.

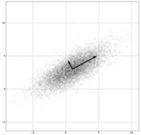
Beispiele:

- Ein achsenparalleles Rechteck der Größe $a \times b$:

$$cov = \frac{1}{12} \cdot \begin{array}{|c|c|} \hline a^2 & 0 \\ \hline 0 & b^2 \\ \hline \end{array}$$

- Ein „diagonaler Strich“ $x = y$ mit $x \in [-a, a]$:

$$cov = \frac{1}{3} \cdot \begin{array}{|c|c|} \hline a^2 & a^2 \\ \hline a^2 & a^2 \\ \hline \end{array}$$



Annahme: die „Richtungen“ kleiner Varianzen entsprechen dem Rauschen und können vernachlässigt werden.

Die Idee: den Merkmalsraum auf einen linearen Unterraum projizieren so, dass die Varianzen im Unterraum so groß wie möglich sind.

Einfachheit halber – die Daten sind bereits zentriert, der Unterraum ist eindimensional, d.h. durch einen Vektor e mit $\|e\|^2 = 1$ angegeben. Projektion eines x auf e ist $\langle x, e\rangle$.

$$\sum_l \langle x^l, e\rangle^2 \to \max_e \qquad \text{s.t. } \|e\|^2 = 1$$

Lagrange Funktion:

$$\sum_l \langle x^l, e\rangle^2 + \lambda\Big(\|e\|^2 - 1\Big) \to \min_\lambda \max_e$$

Ableitung:

$$\sum_l 2\langle x_l, e\rangle \cdot x_l + 2\lambda e = 2e \sum_l x_l \otimes x_l + 2\lambda e = 0$$

$$e \cdot cov = \lambda e$$

$\to$ e ist Eigenvektor der Kovarianzmatrix, λ ist der entsprechende Eigenwert.

Welchen Eigenvektor soll gewählt werden? Die mit einem λ erreichte Varianz ist

$$\sum_l \langle x^l, e\rangle^2 = e \cdot \sum_l x^l \otimes x^l \cdot e = e \cdot cov \cdot e = \|e\|^2 \cdot \lambda = \lambda$$

$\rightarrow$ wähle den Eigenvektor zum größten Eigenwert.

Ähnliche Vorgehensweise: den Merkmalsraum auf einen Unterraum projizieren so, dass die summarische quadratische Abweichung der Datenpunkte von entsprechenden Projektionen so klein wie möglich ist $\rightarrow$ Approximation. Das Ergebnis ist dasselbe.

Allgemein:

1) Berechne die Kovarianzmatrix der Daten $cov = \sum_l x^l \otimes x^l$
2) Suche alle Eigenwerte und Eigenvektoren
3) Ordne sie (fallend) nach Eigenwerten
4) Wähle m Eigenvektoren zu m größten Eigenwerten
5) Die $n \times m$ Projektionsmatrix besteht aus m Spalten, die jeweils die gewählten Eigenvektoren sind.

Weitere Vorgehensweisen

Wann funktioniert PCA nicht?
Wenn die Wahrscheinlichkeitsverteilung nicht Gaussch ist.

Unabhängigkeitsanalyse (ICA)
Transformiere die Daten linear so, dass die Komponenten der transformierte Zufallsgröße so wenig von einander abhängen, wie möglich. Als Maß für Unabhängigkeit:

– Wölbung – die Verteilungen sollen so wenig wie möglich Gaussch sein.
– Entropie

Kernel-PCA
Mache PCA in einem höherdimensionalen Raum, wo „alles linear ist“. Dies entspricht der Projektion der Daten auf einen nicht linearen Unterraum im ursprünglichen Vektorraum.

Literatur:
Jolliffe: Principal Component Analysis. 2002
Aapo Hyvärinen, Juha Karhunen, Erkki Oja: Independent Component Analysis. 2001
`http://www.cs.helsinki.fi/u/ahyvarin/whatisica.shtml`
Kernel-PCA: Bernhard Schölkopf, Alexander Smola, Klaus Robert Müller, 1999–2002