Dimensionsreduktion

Autor und Herausgeber

Alexander Jung*, Department of Computer Science, Aalto University, Espoo, Finland (ORCID 0000-0001-7538-0990)
Editors: Konstantina Olioumtsevits (ORCID 0000-0003-2697-8659), Ekkehard Schnoor (ORCID 0000-0001-8252-1499), Aalto University, Espoo, Finland
*Corresponding author: alex.jung@aalto.fi

Zusammenfassung

Die Dimensionsreduktion verwendet eine Abbildung $\mathbf{\Phi} : \mathbb{R}^d \to \mathbb{R}^{d'}$, um einen ursprünglichen Merkmalsvektor $\mathbf{x} = (x_1, \ldots, x_d)^\top$ durch einen kürzeren Merkmalsvektor $\mathbf{z} = (z_1, \ldots, z_{d'})^\top$ mit $d' < d$ zu ersetzen. Die Verwendung weniger Merkmale verringert die effektive Dimension eines Modells, was die Wahrscheinlichkeit von Überanpassung reduziert. Darüber hinaus verringert sie den Rechenaufwand, der für das Training des Modells und für die Inferenz benötigt wird. Sie ermöglicht außerdem die Visualisierung: Zwei Merkmale können als Koordinaten in einem Streudiagramm verwendet werden. Die Hauptkomponentenanalyse (PCA) erlernt eine lineare Abbildung durch Minimierung des Rekonstruktionsfehlers, während eine zufällige Projektion eine zufällig gezogene lineare Abbildung ganz ohne Training nutzt. Ein Autoencoder erlernt eine nichtlineare Abbildung, die durch ein künstliches neuronales Netz dargestellt wird.

Definition

Ein quadratischer Bildausschnitt der Fotografie in Abb. 2 wird durch 48 Zahlen beschrieben, und zwei daraus berechnete Zahlen genügen, um Himmel von Vegetation zu unterscheiden. Die Dimensionsreduktion verwendet eine Abbildung $\mathbf{\Phi}$, eine Transformation $\mathbf{\Phi} : \mathbb{R}^d \to \mathbb{R}^{d'}$, um den ursprünglichen Merkmalsvektor $\mathbf{x} = (x_1, \ldots, x_d)^\top$ eines Datenpunkts durch einen kürzeren Merkmalsvektor $\mathbf{z} = (z_1, \ldots, z_{d'})^\top$ mit $d' < d$ zu ersetzen (Abb. 1). Die Verwendung eines kürzeren Merkmalsvektors ist in mehrfacher Hinsicht vorteilhaft:

- *Statistischer Vorteil:* Sie reduziert typischerweise die Wahrscheinlichkeit von Überanpassung, da die Verringerung der Anzahl an Merkmale häufig die effektive Dimension eines Modells verringert (Hastie et al., 2009, Abschn. 7.6).
- *Rechnerischer Vorteil:* Die Verwendung weniger Merkmale verringert den Rechenaufwand, der für das Training des Modells und für die Inferenz benötigt wird. Beispielsweise müssen Verfahren der linearen Regression eine Matrix invertieren, deren Größe durch die Anzahl der Merkmale bestimmt wird (Hastie et al., 2009, Gl. (3.6)).
- *Visualisierung:* Es kann eine Abbildung erlernt werden, die zwei Merkmale z_1, z_2 liefert, welche anschließend als Koordinaten in einem Streudiagramm

verwendet werden können. Jeder quadratische Ausschnitt einer Fotografie wird zunächst durch Mittelung auf ein 4×4-Farbraster reduziert, was immer noch zu viele Zahlen zum Plotten sind; Abb. 2 platziert dann jeden Ausschnitt an den beiden Merkmale, die die Hauptkomponentenanalyse (PCA) für ihn erlernt und die den Großteil der Varianz der ursprünglichen Merkmalsvektoren beibehalten (Bishop, 2006, Abschn. 12.1).

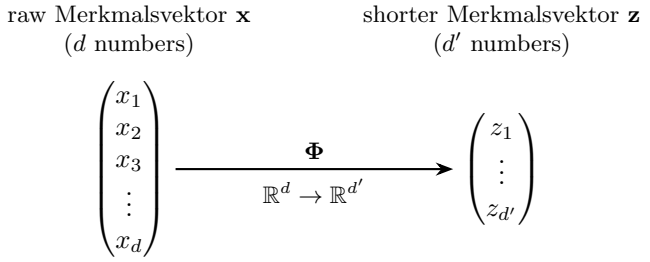

Abbildung 1: Die Dimensionsreduktion erlernt eine Abbildung $\mathbf{\Phi}$, die den ursprünglichen Merkmalsvektor $\mathbf{x}$ eines Datenpunkts, eine Liste von d Zahlen, auf einen kürzeren Merkmalsvektor $\mathbf{z}$ aus $d' < d$ Zahlen abbildet

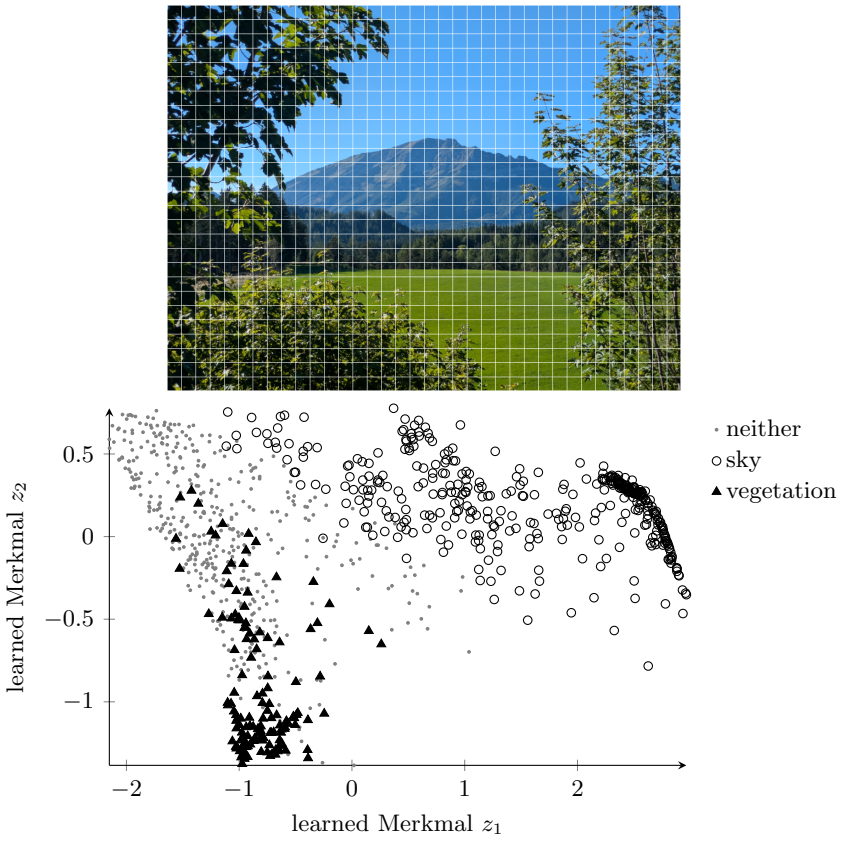

Abbildung 2: Oben: die Fotografie des Ötschermassivs mit dem eingezeichneten Ausschnittraster; jeder 128×128 Pixel große Ausschnitt ist ein Datenpunkt. Unten: jeder Ausschnitt platziert an den beiden Merkmale z_1, z_2, die die PCA für ihn erlernt. Der ursprüngliche Merkmalsvektor eines Ausschnitts ist ein 4×4-Farbraster, also 48 Zahlen; die beiden erlernten Merkmale behalten 86 Prozent seiner Varianz bei. Ein Ausschnitt ist als offener Kreis eingezeichnet, wenn sein Blaukanal die beiden anderen um eine feste Spanne übersteigt (Himmel), und als ausgefülltes Dreieck, wenn dies für seinen Grünkanal gilt (Vegetation); die übrigen Ausschnitte, die Felswand und die gemischten Bereiche entlang der Baumgrenze, sind kleine Punkte. Himmel und Vegetation liegen in getrennten Bereichen, wobei sich 93 Prozent davon auf ihrer jeweils eigenen Seite des Mittelpunkts zwischen den beiden Zentren befinden. Daten erzeugt mit `pythondemos/dimred.py`

Verfahren zur Dimensionsreduktion unterscheiden sich im Hypothesenraum, der für die Abbildung verwendet wird, im Kriterium, nach dem sie erlernt wird, und darin, ob die Abbildung überhaupt aus Daten erlernt wird. Die PCA erlernt eine lineare Abbildung: Die Abbildung ist eine Matrix, die so gewählt

wird, dass der Rekonstruktionsfehler – der quadrierte Abstand zwischen $\mathbf{x}$ und dem aus $\mathbf{z}$ durch eine zweite lineare Abbildung rekonstruierten Merkmalsvektor – im Durchschnitt über die Trainingsdatensatz minimal ist (Bishop, 2006, Abschn. 12.1.2; Hotelling, 1933). Ein Autoencoder erlernt unter demselben Kriterium eine nichtlineare Abbildung, die durch ein künstliches neuronales Netz dargestellt wird (Goodfellow et al., 2016, Abschn. 14.1). Eine zufällige Projektion verwendet eine Zufallsmatrix ganz ohne Training: Nach dem Johnson-Lindenstrauss-Lemma (JL-Lemma) bleiben die Abstände zwischen den Merkmalsvektoren der Datenpunkte näherungsweise erhalten (Shalev-Shwartz and Ben-David, 2014, Lemma 23.4; Vershynin, 2018). Zu den nichtlinearen Verfahren jenseits des Autoencoders zählt die Familie des Manifold-Learning, darunter t-SNE und UMAP, die hauptsächlich zur Visualisierung eingesetzt werden (Ghojogh et al., 2023; Meilă and Zhang, 2024).

Siehe auch: Überanpassung, Autoencoder, zufällige Projektion, PCA, JL-Lemma, Merkmalsvektor, Hypothesenraum.

Querverweise

Literatur

1. Bishop (2006). Pattern Recognition and Machine Learning. Springer Science+Business Media, New York, NY, USA.

2. Ghojogh et al. (2023). Elements of Dimensionality Reduction and Manifold Learning. Springer, Cham.

3. Goodfellow et al. (2016). Deep Learning. MIT Press, Cambridge, MA, USA.

4. Hotelling (1933). Analysis of a Complex of Statistical Variables into Principal Components. Journal of Educational Psychology, 24(6), pp. 417–441.

5. Meilă and Zhang (2024). Manifold learning: what, how, and why. Annu. Rev. Stat. Appl., 11(1), pp. 393–417.

6. Shalev-Shwartz and Ben-David (2014). Understanding Machine Learning: From Theory to Algorithms. Cambridge Univ. Press, New York, NY, USA.

7. Hastie et al. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed., Springer Science+Business Media, New York, NY, USA.

8. Vershynin (2018). High-Dimensional Probability: An Introduction with Applications in Data Science. Cambridge Univ. Press, Cambridge, U.K.