Projektion

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 Projektion eines Vektors auf eine abgeschlossene, nichtleere Teilmenge eines Euclidean space ist ein Punkt in der Teilmenge, der dem Vektor in der Euclidean norm am nächsten liegt. Ist die Teilmenge konvex, so ist dieser nächste Punkt eindeutig und die Lösung eines konvexen Optimierungsproblems. Ist die Teilmenge ein Unterraum, so ist die Projektion eine lineare Abbildung, nämlich die orthogonale Projektion auf den Unterraum. Jede Projektion ist idempotent, und eine lineare Abbildung eines Hilbertraum in sich selbst ist genau dann eine orthogonale Projektion, wenn die Abbildung idempotent und selbstadjungiert ist. Im machine learning (ML) erzwingen Projektionen Einschränkungen für Modellparameter während des Trainings: projected gradient descent (projected GD) wechselt Gradientenschritte mit Projektionen auf die Menge der Nebenbedingungen ab, z. B. die ℓ_1-Kugel in einer äquivalenten Formulierung des least absolute shrinkage and selection operator (Lasso).

Definition

Das least absolute shrinkage and selection operator (Lasso) lernt ein lineares Modell, dessen Modellparameter $\mathbf{w}$ eine ℓ_1-Norm unterhalb einer vorgegebenen Schranke halten müssen, und ein Gradientenschritt kann $\mathbf{w}$ außerhalb dieser Menge bewegen. Das Ergebnis muss dann durch einen nächstgelegenen zulässigen Vektor ersetzt werden, und das Training jedes machine learning (ML) Modells unter einer Nebenbedingungsmenge sieht sich bei jedem Schritt mit demselben Ersatz konfrontiert. Die Projektion formalisiert diesen Ersatz.

Man betrachte eine nichtleere closed set $\mathcal{W} \subseteq \mathbb{R}^d$ im d-dimensionalen Euclidean space. Die Projektion $P_{\mathcal{W}}\big(\mathbf{w}\big)$ eines Vektors $\mathbf{w} \in \mathbb{R}^d$ auf $\mathcal{W}$ ist

$$P_{\mathcal{W}}\big(\mathbf{w}\big) = \operatorname*{arg\,min}_{\mathbf{w}' \in \mathcal{W}} \|\mathbf{w} - \mathbf{w}'\|_2,$$

ein Vektor in $\mathcal{W}$, der $\mathbf{w}$ in der Euclidean norm am nächsten liegt. Das minimum existiert für jede nichtleere closed set $\mathcal{W}$ (Boyd and Vandenberghe, 2004, Sect. 8.1), es können jedoch mehrere nächstgelegene Vektoren existieren. Für die Zwei-Punkt-Menge $\mathcal{W} = \big\{(-1,0)^\top, (1,0)^\top\big\}$ sind beide Elemente am nächsten zu $(0, 0.7)^\top$, und ebenso zu jedem anderen Punkt auf der vertikalen Achse, die

genau aus den Punkten besteht, die von den beiden Elementen gleich weit entfernt sind. Ist $\mathcal{W}$ abgeschlossen und konvex, so ist der nächstgelegene Punkt in der Euclidean norm eindeutig (Boyd and Vandenberghe, 2004, Sect. 8.1).

Die Konstruktion erstreckt sich über den Euclidean space hinaus: Eine nichtleere, abgeschlossene, konvexe Teilmenge eines Hilbertraum enthält für jeden Punkt des Raumes einen eindeutigen nächstgelegenen Punkt (Bauschke and Combettes, 2011, Thm. 3.14). Ist $\mathcal{W}$ ein Unterraum, d.h. eine Teilmenge, die alle Summen und skalaren Vielfachen ihrer Elemente enthält (wie eine Gerade durch den Ursprung), so ist die Abbildung $\mathbf{w} \mapsto P_{\mathcal{W}}(\mathbf{w})$ eine lineare Abbildung: Sie ist die orthogonale Projektion auf $\mathcal{W}$ (Axler, 2024, Thm. 6.57 and Thm. 6.61; Bauschke and Combettes, 2011, Cor. 3.22).

Jede Projektion ist idempotent: Der Vektor $P_{\mathcal{W}}(\mathbf{w})$ liegt in $\mathcal{W}$, so ist sein nächstgelegener Punkt in $\mathcal{W}$ er selbst und $P_{\mathcal{W}}\big(P_{\mathcal{W}}(\mathbf{w})\big) = P_{\mathcal{W}}(\mathbf{w})$. Eine lineare Abbildung P eines Hilbertraum in sich selbst ist genau dann die orthogonale Projektion auf den Bildbereich von P, wenn P idempotent und selbstadjungiert ist, d.h. $\langle P\mathbf{u}, \mathbf{v}\rangle = \langle \mathbf{u}, P\mathbf{v}\rangle$ für alle $\mathbf{u}, \mathbf{v}$ (Axler, 2024, Sect. 6C, Exercise 9, and Sect. 7A, Exercise 20; Rudin, 1991, Thm. 12.14). Für eine Matrix $\mathbf{P} \in \mathbb{R}^{d\times d}$, die auf dem Euclidean space operiert, ist Idempotenz die Bedingung $\mathbf{P}^2 = \mathbf{P}$ und Selbstadjungiertheit die Bedingung $\mathbf{P}^\top = \mathbf{P}$. Idempotenz allein ist nicht ausreichend, damit eine lineare Abbildung eine orthogonale Projektion ist: Eine idempotente lineare Abbildung, die nicht selbstadjungiert ist, ist eine schiefe Projektion (Meyer, 2000, Sect. 5.9). Zum Beispiel erfüllt die Matrix $\mathbf{P} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$ auf $\mathbb{R}^2$ die Bedingung $\mathbf{P}^2 = \mathbf{P}$, aber $\mathbf{P}^\top \neq \mathbf{P}$. Ihr Bildbereich ist die horizontale Achse, und sie bildet $\mathbf{w} = (0, 1)^\top$ auf $\mathbf{P}\mathbf{w} = (1, 0)^\top$ ab, während der Punkt der Achse, der $\mathbf{w}$ am nächsten liegt, der Ursprung $(0, 0)^\top$ ist.

Im ML erzwingen Projektionen Einschränkungen für Modellparameter während des Trainings. Eine lineare Regression mit der Nebenbedingung $\|\mathbf{w}\|_1 \leq \tau$, für eine vorgegebene Konstante $\tau > 0$, ist eine äquivalente Formulierung des Lasso, für einen Wert ihres Strafterms Hyperparameters, der von τ und dem Trainingsdatensatz abhängt (Hastie et al., 2009, Sect. 3.4.2; Tibshirani, 1996). Dieses restringierte Problem kann durch projected gradient descent (projected GD) gelöst werden (Beck, 2017, Sect. 10.2). Jede Iteration führt einen Gradientenschritt auf dem Trainingsfehler aus und projiziert dann das Ergebnis auf die ℓ_1-Kugel $\mathcal{W} = \{\mathbf{w}' \in \mathbb{R}^d : \|\mathbf{w}'\|_1 \leq \tau\}$. Der Projektionsschritt erzwingt die Nebenbedingung: Ohne ihn kann ein grundlegendes Gradientenabstieg auf dem Trainingsfehler Iterationen mit $\|\mathbf{w}\|_1 > \tau$ erzeugen, und das Ergebnis löst das unrestringierte Problem anstelle des restringierten. Da $\mathcal{W}$ kein Unterraum ist (das Skalieren des Vektors $(\tau, 0)^\top \in \mathcal{W}$ mit 2 ergibt einen Vektor außerhalb der Kugel), ist diese Projektion keine lineare Abbildung, sondern läuft auf die Lösung eines konvexen Optimierungsproblems hinaus (Boyd and Vandenberghe, 2004, Sect. 8.1.1) (siehe Abb. 1). Eine solche Projektion unterscheidet sich vom proximaler Operator: Der proximaler Operator minimiert über ganz $\mathbb{R}^d$ die Summe einer konvexen Funktion $f(\mathbf{w}')$ und der Hälfte des quadrierten euklidischer Abstand zu $\mathbf{w}$, während die Projektion die Abstand allein über der

gegebenen Menge $\mathcal{W}$ minimiert.

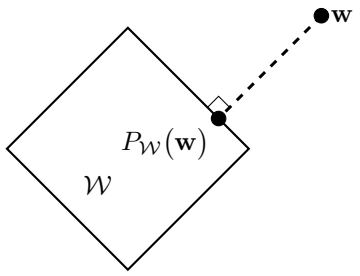

Abbildung 1: Die Projektion $P_{\mathcal{W}}(\mathbf{w})$ ist der Punkt der ℓ_1-Kugel $\mathcal{W} = \{\mathbf{w}' \in \mathbb{R}^2 : \|\mathbf{w}'\|_1 \leq \tau\}$, der einem Vektor $\mathbf{w} \in \mathbb{R}^2$ in der Euclidean norm am nächsten liegt; das kleine Quadrat markiert den rechten Winkel zwischen dem Segment von $\mathbf{w}$ zu $P_{\mathcal{W}}(\mathbf{w})$ und dem Rand der Kugel, der $P_{\mathcal{W}}(\mathbf{w})$ enthält.

Siehe auch: Euclidean space, Vektor, minimum, konvex, orthogonale Projektion, proximaler Operator, projected GD.

Querverweise

Literatur

1. Axler (2024). Linear Algebra Done Right. 4th ed., Springer, Cham, Switzerland.
2. Bauschke and Combettes (2011). Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer Science+Business Media, New York, NY, USA.
3. Beck (2017). First-Order Methods in Optimization. SIAM-Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.
4. Boyd and Vandenberghe (2004). Convex Optimization. Cambridge Univ. Press, Cambridge, U.K.
5. Meyer (2000). Matrix Analysis and Applied Linear Algebra. SIAM, Philadelphia, PA, USA.
6. Rudin (1991). Functional Analysis. 2nd ed., McGraw-Hill, New York, NY, USA.
7. Tibshirani (1996). Regression Shrinkage and Selection via the Lasso. J. Roy. Statist. Soc.: Ser. B (Methodological), 58(1), pp. 267–288.
8. Hastie et al. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed., Springer Science+Business Media, New York, NY, USA.