

Lineare Klassifikation

Anne Driemel

Letzte Aktualisierung: 14. Mai 2020

Die lineare Klassifikation ist eine der grundlegendsten Methoden im Maschinellen Lernen. Die entsprechende Hypothesenklasse \mathcal{H} ist definiert als die Menge von Funktionen der Form $h_{a,b} : \mathbb{R}^d \rightarrow \{-1, +1\}$ mit $a \in \mathbb{R}^d, b \in \mathbb{R}$ und¹

$$h_{a,b}(x) = \begin{cases} +1 & \text{falls } \langle a, x \rangle \geq b \\ -1 & \text{sonst} \end{cases}$$

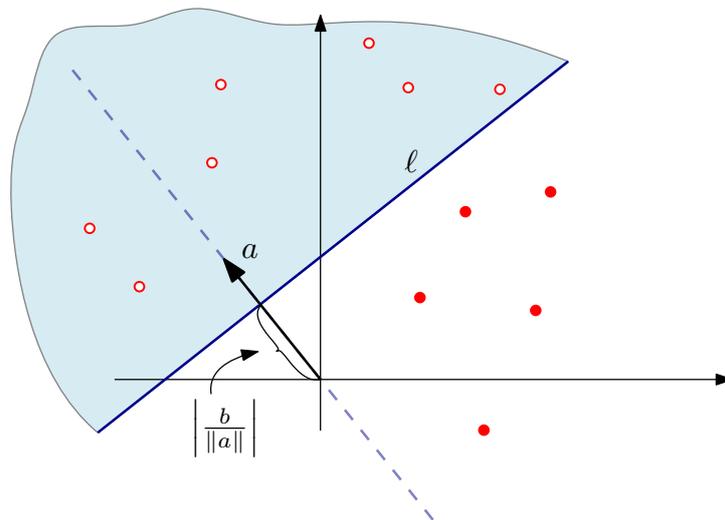
Wir können dies wieder äquivalent als Mengensystem beschreiben mit den Mengen

$$r_{a,b} = \left\{ x \in \mathbb{R}^d \mid \langle a, x \rangle \geq b \right\}$$

Die Menge $r_{a,b}$ definiert einen Halbraum von \mathbb{R}^d . Ein Halbraum ist eine Menge, die durch eine Hyperebene beschränkt ist. In unserem Fall ist das die folgende Hyperebene

$$\ell = \left\{ x \in \mathbb{R}^d \mid \langle a, x \rangle = b \right\}$$

Im \mathbb{R}^2 können wir uns das geometrisch vorstellen. Die Hyperebene ℓ ist orthogonal zu der Geraden g durch den Nullpunkt, die den Vektor a enthält und schneidet diese Gerade im Abstand $\left| \frac{b}{\|a\|} \right|$ zum Nullpunkt². Der Halbraum $r_{a,b}$ umfasst alle Punkte zu der Seite von ℓ die durch die Richtung des Vektors a angegeben ist.



¹Für Vektoren $x = (x_1, \dots, x_d)$ und $y = (y_1, \dots, y_d)$ ist das Skalarprodukt definiert als $\langle x, y \rangle = \sum_{i=1}^d x_i y_i$

²Die Norm eines Vektors $x = (x_1, \dots, x_d)$ ist hier definiert als $\|x\| = \sqrt{\sum_{i=1}^d x_i^2}$

1 VC-dimension von Halbräumen

Wir wollen heute die VC-dimension des Mengensystems aller Halbräume analysieren. Zunächst wollen wir dazu ein paar grundlegende Begriffe einführen.

Definition 6.1 (Affinkombination). Für beliebige Punkte $p_1, \dots, p_n \in \mathbb{R}^d$ und Parameter $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ mit $\sum_{i=1}^n \alpha_i = 1$ ist $\sum_{i=1}^n \alpha_i p_i$ eine Affinkombination. Die Punkte p_1, \dots, p_n sind affin abhängig wenn es einen Punkt p_i gibt, sodass $p_i = \sum_{\substack{j=1 \\ j \neq i}}^n \alpha_j p_j$. Also genau dann wenn wir einen Punkt der Menge durch eine Affinkombination der anderen Punkte ausdrücken können. Die Menge aller Affinkombinationen einer festen Menge wird als ihre affine Hülle bezeichnet.

Eine Affinkombination ist also eine Linearkombination mit der zusätzlichen Bedingung, dass die Summe der Koeffizienten 1 ergibt. Mithilfe von Affinkombinationen lassen sich Geraden, Ebenen und Hyperebenen darstellen.

Beispiel 6.2. Die Menge aller Affinkombinationen zweier Punkte $p_1, p_2 \in \mathbb{R}^2$ ist

$$\{ tp_1 + (1-t)p_2 \mid t \in \mathbb{R} \}$$

Das ist die Menge aller Punkte auf der Geraden, welche p_1 und p_2 enthält.

Lemma 6.3. Jede Menge von $d+2$ Punkten im \mathbb{R}^d ist affin abhängig.

Beweis. Sei $A = \{p_1, \dots, p_{d+2}\}$. Setze ein beliebiges p_i fest und betrachte für $i \neq j$ die $d+1$ Punkte $q_j = p_j - p_i$. Da $d+1$ Punkte linear abhängig sind, existieren Parameter β_j für $i \neq j$ und ein Punkt q_r sodass

$$q_r = \sum_{\substack{j=1 \\ i \neq j \text{ und } j \neq r}}^{d+2} \beta_j q_j.$$

Somit gilt

$$p_r = q_r + p_i = \left(\sum_{\substack{j=1 \\ i \neq j \text{ und } j \neq r}}^{d+2} \beta_j q_j \right) + p_i = \sum_{\substack{j=1 \\ i \neq j \text{ und } j \neq r}}^{d+2} \beta_j p_j - \left(\sum_{\substack{j=1 \\ i \neq j \text{ und } j \neq r}}^{d+2} \beta_j \right) p_i + p_i$$

Jetzt können wir $\beta_i = - \left(\sum_{\substack{j=1 \\ i \neq j \text{ und } j \neq r}}^{d+2} \beta_j \right) + 1$ definieren und somit haben wir $\sum_{\substack{j=1 \\ j \neq r}}^{d+2} \beta_j = 1$ womit die Bedingung für eine Affinkombination erfüllt ist. \square

Definition 6.4 (Konvexkombination). Für beliebige Punkte $p_1, \dots, p_n \in \mathbb{R}^d$ und Parameter $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ mit $\sum_{i=1}^n \alpha_i = 1$ und $\alpha_i \geq 0$ für alle $1 \leq i \leq n$ ist $\sum_{i=1}^n \alpha_i p_i$ eine Konvexkombination. Die Menge aller Konvexkombinationen einer festen Menge wird als ihre konvexe Hülle bezeichnet.

Beispiel 6.5. Die konvexe Hülle von zwei Punkten $p_1, p_2 \in \mathbb{R}^2$ ist die Menge

$$\{ tp_1 + (1-t)p_2 \mid t \in [0, 1] \}.$$

Das ist die Strecke mit Endpunkten p_1 und p_2 .

Beispiel 6.6. Die konvexe Hülle von drei Punkten $p_1, p_2, p_3 \in \mathbb{R}^2$ ist die Menge

$$\{ \alpha_1 p_1 + \alpha_2 p_2 + \alpha_3 p_3 \mid \alpha_1, \alpha_2, \alpha_3 \geq 0 \text{ und } \alpha_1 + \alpha_2 + \alpha_3 = 1 \}.$$

Das ist das Dreieck mit den Eckpunkten p_1, p_2 und p_3 .

1.1 Radon's Lemma

Lemma 6.7 (Radon's Lemma). *Für jede Menge A von $d+2$ Punkten im \mathbb{R}^d existieren Teilmengen $A_1, A_2 \subseteq A$ mit $A_1 \cap A_2 = \emptyset$ und ein Punkt q , sodass q sowohl als Konvexkombination von A_1 dargestellt werden kann, als auch eine Konvexkombination von A_2 . Wir bezeichnen q als Radonpunkt der Mengen A_1 und A_2 .*

Beweis. Sei $A = \{p_1, \dots, p_{d+2}\}$. Da wir $d+2$ Punkte haben, sind diese affin abhängig. Das heißt, es gibt ein $p_i \in A$ welches durch eine Affinkombination der anderen Punkte in A dargestellt werden kann. Also existieren Parameter α_j für $1 \leq j \leq n$ mit $i \neq j$, sodass

$$p_i = \sum_{\substack{j=1 \\ i \neq j}}^{d+2} \alpha_j p_j \quad \text{mit} \quad \sum_{\substack{j=1 \\ i \neq j}}^{d+2} \alpha_j = 1$$

Setzen wir nun $\alpha_i = -1$, dann können wir $\alpha_i p_i$ auf beiden Seiten der Gleichung addieren und bekommen

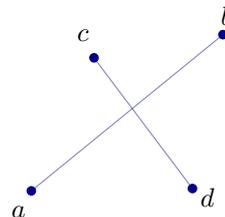
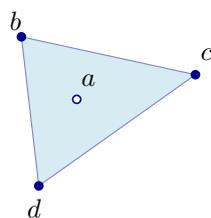
$$0 = \sum_{j=1}^{d+2} \alpha_j p_j \quad \text{mit} \quad \sum_{j=1}^{d+2} \alpha_j = 0.$$

Wir definieren nun zwei Indexmengen $I_1 = \{i \mid \alpha_i > 0\}$ und $I_2 = \{i \mid \alpha_i < 0\}$. Durch äquivalente Umformung bekommen wir

$$-\sum_{i \in I_2} \alpha_i p_i = \sum_{i \in I_1} \alpha_i p_i \quad \text{und} \quad -\sum_{i \in I_2} \alpha_i = \sum_{i \in I_1} \alpha_i$$

Setzen wir nun $\gamma = \sum_{i \in I_1} \alpha_i$, dann definiert $q_1 = \sum_{i \in I_1} \beta_i p_i$ mit $\beta_i = \frac{\alpha_i}{\gamma}$ eine Konvexkombination der Punkte in A mit Index in I_1 . Ähnlich definiert $q_2 = \sum_{i \in I_2} \beta_i p_i$ mit $\beta_i = -\frac{\alpha_i}{\gamma}$ eine Konvexkombination der Punkte in A mit Index in I_2 . Weiter ist $q_1 = q_2$ und $I_1 \cap I_2 = \emptyset$. Damit ist der Satz bewiesen. \square

Beispiel 6.8 (Radonpunkt). *Für 4 verschiedene Punkte a, b, c, d in der Ebene gibt es im Prinzip zwei Möglichkeiten wie die Teilmengen in Radon's Lemma zueinander liegen können.*



- (a) $A_1 = \{a\}, A_2 = \{b, c, d\}$.

Ein Punkt a ist in der konvexen Hülle der anderen Punkte $\{b, c, d\}$ enthalten. Hier ist a ein Radonpunkt.

- (b) $A_1 = \{a, b\}, A_2 = \{c, d\}$.

Zwei Strecken \overline{ab} und \overline{cd} schneiden sich in einem Punkt. Der Schnittpunkt stellt einen Radonpunkt dar.

1.2 Beweis der VC-dimension

Satz 6.9. Die VC-dimension von Halbräumen in \mathbb{R}^d ist höchstens $d + 1$.

Beweis. Sei \mathcal{R} das Mengensystem von Halbräumen in \mathbb{R}^d . Das heißt, jede Menge in \mathcal{R} ist von der Form $r_{a,b} = \{ x \in \mathbb{R}^d \mid \langle a, x \rangle \geq b \}$ mit $a \in \mathbb{R}^d, b \in \mathbb{R}$. Wir zeigen, dass die VC-dimension höchstens $d + 1$ sein kann. Angenommen dem wäre nicht so. Wir führen diese Annahme zu einem Widerspruch. Sei $A = \{p_1, \dots, p_{d+2}\} \subseteq \mathbb{R}^d$ eine Menge von $d + 2$ Punkten die durch \mathcal{R} aufgespalten wird. Laut Radon's Lemma gibt es zwei disjunkte Teilmengen $A_1, A_2 \subseteq A$ die einen gemeinsamen Radonpunkt q besitzen. Das heißt, es gibt Konvexkombinationen

$$q = \sum_{i \in I_1} \alpha_i p_i \quad \text{und} \quad q = \sum_{i \in I_2} \beta_i p_i$$

wobei I_1 und I_2 die Indexmengen von A_1 und A_2 sind.

Sei $r_{a,b}$ der Halbraum, der A_1 von A abspaltet. Dann ist

$$A_1 = r_{a,b} \cap A \quad \text{und} \quad A_2 \cap r_{a,b} = \emptyset$$

also ist

$$\langle a, p_i \rangle \geq b \quad \text{für alle } i \in I_1$$

und

$$\langle a, p_i \rangle < b \quad \text{für alle } i \in I_2$$

Betrachten wir nun $\langle a, q \rangle$, so können wir mit der ersten Konvexkombination unter Nutzung der Linearität des Skalarprodukts herleiten

$$\langle a, q \rangle = \left\langle a, \sum_{i \in I_1} \alpha_i p_i \right\rangle = \sum_{i \in I_1} \alpha_i \langle a, p_i \rangle \geq \sum_{i \in I_1} \alpha_i b = b$$

Die Abschätzung $\alpha_i \langle a, p_i \rangle \leq \alpha_i b$ können wir natürlich nur machen, da alle α_i positiv sind. Die letzte Gleichheit folgt aus der anderen Bedingung an die Konvexkombination, dass die Summe der Koeffizienten gleich 1 ist. Ähnlich können wir mit der zweiten Konvexkombination herleiten

$$\langle a, q \rangle = \left\langle a, \sum_{i \in I_2} \beta_i p_i \right\rangle = \sum_{i \in I_2} \beta_i \langle a, p_i \rangle < \sum_{i \in I_2} \beta_i b = b$$

Damit ergibt sich ein Widerspruch mit $b > \langle a, q \rangle \geq b$. □

Geometrisch kann man sich den Beweis für die obere Schranke wie folgt veranschaulichen. Laut Radon's Lemma gibt es zwei disjunkte Teilmengen $A_1, A_2 \subseteq A$, sodass die beiden konvexen Hüllen von A_1 und A_2 nicht zueinander disjunkt sind. Ein Halbraum, der A_1 von A abspaltet ist durch eine Hyperebene beschränkt, die A_1 und A_2 linear separiert. Die Hyperebene würde dann aber auch die beiden konvexen Hüllen linear separieren.

Satz 6.10. Die VC-dimension von Halbräumen in \mathbb{R}^d ist mindestens $d + 1$.

Beweis. Wir zeigen, dass es eine Punktmenge mit $d + 1$ Punkten gibt, die durch \mathcal{R} aufgespalten wird. Dafür konstruieren wir eine Menge $A \subseteq \mathbb{R}^d$ mit $|A| = d + 1$, die durch \mathcal{R} aufgespalten wird. Sei e_i der Einheitsvektor, der überall 0 ist und nur an der i ten Koordinate eine 1 hat. Wir definieren als Menge A die d Einheitsvektoren und den Nullvektor $e_0 = (0, \dots, 0)$. Nun können

wir für jede Teilmenge $A' \subseteq A$ zeigen, dass A' abgespalten wird. Wir wählen $a = (a_1, \dots, a_d) \in \mathbb{R}^d$ und $b \in \mathbb{R}$ wie folgt

$$a_i = \begin{cases} 1 & \text{falls } e_i \in A', \\ -1 & \text{sonst} \end{cases} \quad \text{und } b = \begin{cases} 0 & \text{falls } e_0 \in A' \\ 1 & \text{sonst} \end{cases}$$

Dann gilt $\langle a, e_0 \rangle = 0$ und für $0 < i \leq d$

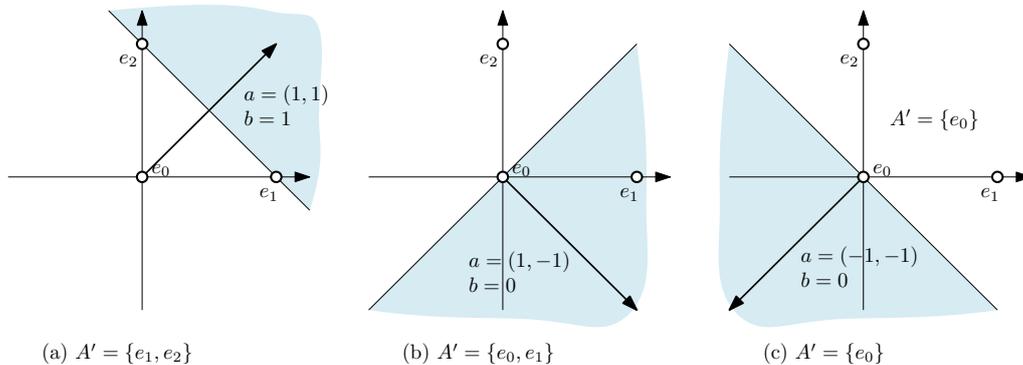
$$\langle a, e_i \rangle = \begin{cases} 1 & \text{falls } e_i \in A' \\ -1 & \text{sonst} \end{cases}$$

Durch eine Fallanalyse kann man nun zeigen, dass für alle $e_i \in A$ gilt, dass

$$\langle a, e_i \rangle \geq b \iff e_i \in A'$$

Somit kann die Menge A' immer durch einen Halbraum abgespalten werden. □

Beispiel 6.11. *Im Beispiel von $d = 2$ können wir exemplarisch Teilmengen von $A = \{e_0, e_1, e_2\}$ aus dem obigen Beweis und die zugehörigen Halbräume visualisieren.*



2 Homogene Halbräume

Sei \mathcal{R}_0 das Mengensystem aller Halbräume der Form

$$r_w = \left\{ x \in \mathbb{R}^d \mid \langle w, x \rangle \geq 0 \right\} \quad \text{mit } w \in \mathbb{R}^d$$

Wir bezeichnen Halbräume dieser Form als homogene Halbräume.

Mit dem Mengensystem \mathcal{R}_0 sind auch allgemeine Halbräume im \mathbb{R}^{d-1} darstellbar. Insbesondere können wir eine Funktion $\phi : \mathbb{R}^{d-1} \rightarrow \mathbb{R}^d$ definieren als $\phi(x) = (x_1, \dots, x_{d-1}, 1)$ und dann existiert für jeden Halbraum

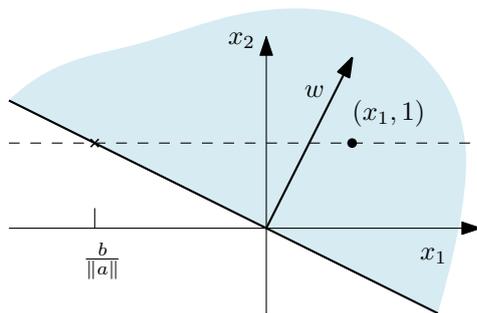
$$r_{a,b} = \left\{ x \in \mathbb{R}^{d-1} \mid \langle a, x \rangle \geq b \right\} \quad \text{mit } a \in \mathbb{R}^{d-1}, b \in \mathbb{R}$$

ein Halbraum $r_w \in \mathcal{R}_0$ sodass für alle $x \in \mathbb{R}^{d-1}$ gilt

$$x \in r_{a,b} \iff \phi(x) \in r_w$$

insbesondere können wir $w = (a_1, \dots, a_{d-1}, -b)$ wählen damit dies erfüllt ist.

Beispiel 6.12. *Für $d = 2$ können wir uns diese Halbräume im konkreten Beispiel wie folgt veranschaulichen. Sei $a = 1, b = -2$, dann ist $w = (1, 2)$. Die Gerade, die den Halbraum r_w im \mathbb{R}^2 beschränkt schneidet die horizontale Gerade bei $y = 1$ in der x -Koordinate $\frac{b}{\|a\|}$. Die Punkte auf der Horizontalen, die in r_w enthalten sind, entsprechen den Punkten in \mathbb{R} , die in $r_{a,b} \subseteq \mathbb{R}$ enthalten sind.*



Referenzen

- Foundations of Machine Learning, Kapitel 3.3. (insbesondere Beispiel 3.12).
- Understanding Machine Learning, Kapitel 9 (insbesondere Kapitel 9.1.3).
- Jiří Matoušek, Lectures on Discrete Geometry, Springer Graduate Texts in Mathematics.