Hypothese

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

Eine Hypothese ist eine Abbildung $h : \mathcal{X} \to \mathcal{Y}$ vom Merkmalsraum $\mathcal{X}$ in den Labelraum $\mathcal{Y}$. Für einen gegebenen Datenpunkt mit Merkmalen $\mathbf{x}$ liefert die Hypothese eine Vorhersage $h(\mathbf{x})$ für das Label y. Eine Methode des maschinellen Lernens wählt eine einzelne gelernte Hypothese $\widehat{h}$ aus einem eingeschränkten Hypothesenraum $\mathcal{H} \subseteq \mathcal{Y}^{\mathcal{X}}$ aus. Diese Einschränkung spiegelt die endlichen Rechenressourcen einer praktischen Methode des maschinellen Lernens wider. Eine auf empirischer Risikominimierung basierende Methode findet $\widehat{h}$ durch eine Suche, welche die Struktur von $\mathcal{H}$ nutzt, z. B. mittels Gradientenabstieg auf einem parametrischen Modell.

Definition

Die Vorhersage der morgigen Lufttemperatur um 15:00 Uhr (UTC) aus dem heutigen Messwert um 06:00 Uhr ist eine typische Aufgabe des maschinellen Lernens. Eine Hypothese ist die Abbildung, die diese Vorhersage durchführt. Formal ist eine Hypothese eine Abbildung (auch Funktion genannt) $h : \mathcal{X} \to \mathcal{Y}$ vom Merkmalsraum $\mathcal{X}$ in den Labelraum $\mathcal{Y}$ (Vapnik, 2000, Sect. 1.1; Shalev-Shwartz and Ben-David, 2014, Sect. 2.1). Für einen gegebenen Datenpunkt mit Merkmalen $\mathbf{x}$ liefert die Hypothese h eine Schätzung $\widehat{y} = h(\mathbf{x})$ des wahren Labels y. Dieselbe Struktur – Merkmale als Eingabe und Vorhersage als Ausgabe – tritt in einer ganz anderen Anwendung auf: Abb. 1 zeigt eine Hypothese, die einen Freddie-Ähnlichkeitswert aus den Stichprobenn einer Audioaufnahme vorhersagt, d. h. aus den gemessenen Amplituden zu aufeinanderfolgenden Zeitpunkten.

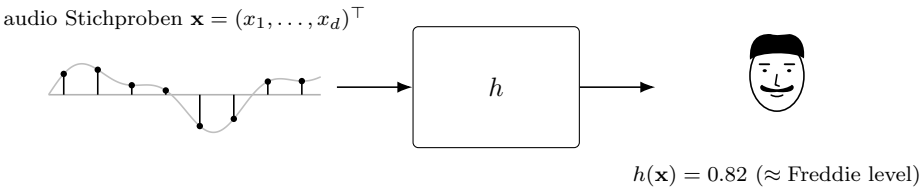


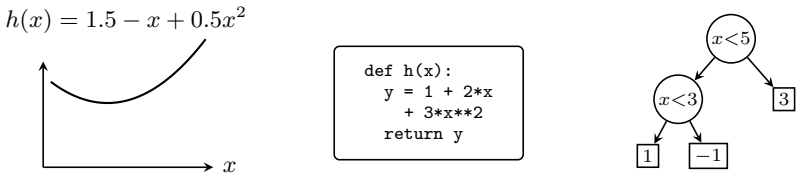
Abbildung 1: Hypothese $h : \mathcal{X} \to \mathcal{Y}$, welche die Merkmale $\mathbf{x} \in \mathcal{X}$ eines Datenpunkts auf eine Vorhersage $h(\mathbf{x}) \in \mathcal{Y}$ des Labels abbildet. Hier verwendet die Anwendung des maschinellen Lernens unter `https://freddiemeter.withyoutube.com/` die Stichproben einer Audioaufnahme als Merkmale, um vorherzusagen, wie stark der Gesang einer Person dem von Freddie Mercury ähnelt. Der angezeigte Wert $h(\mathbf{x}) = 0.82$ gibt eine Ähnlichkeit von 82% an

Eine Methode des maschinellen Lernens zielt darauf ab, eine Hypothese h zu finden (oder zu lernen), sodass $y \approx h(\mathbf{x})$ für jeden Datenpunkt mit Merkmalen $\mathbf{x}$ und Label y gilt. Endliche Rechenressourcen erfordern eine Einschränkung der Suche auf eine Teilmenge von

$$\mathcal{Y}^{\mathcal{X}} = \{h : \mathcal{X} \to \mathcal{Y}\},$$

der Menge aller Funktionen von $\mathcal{X}$ nach $\mathcal{Y}$. Diese Teilmenge ist der der Methode zugrunde liegende Hypothesenraum $\mathcal{H}$ (auch Modell genannt). Die Ausgabe eines Algorithmus des maschinellen Lernens, der auf einem Trainingsdatensatz arbeitet, ist eine einzelne gelernte Hypothese $\widehat{h} \in \mathcal{H}$, die mittels empirischer Risikominimierung gewonnen wird, d. h. ein Element aus $\mathcal{H}$, welches das empirische Risiko auf dem Trainingsdatensatz minimiert (Shalev-Shwartz and Ben-David, 2014, Sect. 2.2). Die Suche prüft die Hypothesen nicht einzeln nacheinander. Bei einem parametrischen Modell, dessen Hypothesen durch einen Vektor von Gewichten indiziert sind, bewegt sich der Gradientenabstieg durch $\mathcal{H}$, indem er dem Gradienten des empirischen Risikos folgt, was nur möglich ist, weil $\mathcal{H}$ diese Struktur aufweist.

Verschiedene Hypothesenräume lassen unterschiedliche Darstellungen ihrer Hypothesenabbildungen zu. Die polynomielle Regression (mit einem einzigen skalaren Merkmal x) stellt jedes $h \in \mathcal{H}$ als algebraischen Ausdruck $h(x) = \sum_{j=0}^{p} w_j x^j$ dar, wobei p der Grad des Polynoms ist und die Koeffizienten w_j die Modellparameter sind. Ein Hypothesenraum von Computerprogrammen stellt h als ausführbaren Quellcode dar, z. B. als eine Python-Funktion mit vorgeschriebener Signatur. Ein Entscheidungsbaum oder ein künstliches neuronales Netz stellt h als Signalflussdiagramm dar, dessen Knoten einfache Operationen anwenden (z. B. Vergleiche, gewichtete Summen, nichtlineare Aktivierungen) und dessen Kanten die Reihenfolge festlegen, in der diese Operationen verkettet werden. Diese Darstellungen unterscheiden sich darin, welche Funktionen sie ausdrücken können und wie effizient das Problem der empirischen Risikominimierung über dem entsprechenden $\mathcal{H}$ gelöst werden kann. Abb. 2 zeigt die drei Darstellungen nebeneinander.



(a) (b) (c)

Abbildung 2: Drei Darstellungen einer Hypothesenabbildung. (a) Als algebraischer Ausdruck, hier ein Polynom zweiten Grades in einem skalaren Merkmal x. (b) Als Quellcode, hier eine Python-Funktion. (c) Als Signalflussdiagramm, hier ein kleiner Entscheidungsbaum, der einen Datenpunkt basierend auf Schwellenwertvergleichen für das Merkmal x zu einer Vorhersage von 3, 1 oder -1 leitet

Synonyme: Prädiktor.

Siehe auch: Funktion, Vorhersage, Prädiktor, Hypothesenraum, Datenpunkt, Label, ERM.

Querverweise

Literatur

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

2. Vapnik (2000). The Nature of Statistical Learning Theory. 2nd ed., Springer, New York, NY, USA.