

WSI für Informatik an der Karls-Eberhardt Universität Tübingen

Machine Learning

Übungsblatt 7

Lea Bey - Benjamin Çoban - Thomas Stüber

13. Juli 2017

7

Aufgabe 7.2.

1. Betrachten wir folgende Gleichung:

$$p(x) \approx \frac{\frac{k}{n}}{V}$$

Nun müssen wir erst einmal betrachten, was die einzelnen Komponenten beschreiben:

- Die Wahrscheinlichkeit, dass ein Vektor x in eine Region \mathcal{R} liegt, wird beschrieben durch P .

$$P = \int_{\mathcal{R}} p(x') dx'$$

- Weiter nehmen wir an, dass n Proben x_1, \dots, x_n gleichverteilt vorliegen bzgl $p(x)$. So ist die Wahrscheinlichkeit binomialverteilt und es gilt für k Elemente in \mathcal{R} aus n Elementen:

$$P_k = \binom{n}{k} \cdot P^k (1 - P)^{n-k}$$

- Der Erwartungswert von k ist demnach $E(k) = n \cdot P$.
- Um nun die Dichte um eine Probe x zu schätzen, betrachten wir eine Regionenfolge \mathcal{R}_n , welche k_n Punkte enthält - inklusive x . Weiter ist V_n die Fläche von \mathcal{R}_n . Durch steigendes n werden dann doch relativ wenige Punkte in \mathcal{R}_n fallen. Somit ist das Verhältnis zwischen P und V relativ gleichbleibend.
- Betrachten wir nun den Grenzwert von

$$\lim_{n \rightarrow \infty} p_n(x) = \frac{\frac{k_n}{n}}{V_n}$$

Somit konvergiert der Schätzer $p_n(x)$ gegen das angestrebte Verhältnis zwischen P und V und entspricht deshalb $p(x)$.

$$p_n(x) = \frac{\int_{\mathcal{R}} p(x') dx'}{V_n} \xrightarrow[n \rightarrow \infty]{} \frac{\frac{k}{n}}{V}$$

2. Die Strategie bei beiden Modellen:

- Beim Parzenfenster-Verfahren wird die Initialregion V in Abhängigkeit von n begrenzt, zum Beispiel $V_n = \frac{1}{\sqrt{n}}$. Wir gehen davon aus, dass die Region \mathcal{R}_n temporär ein d -dimensionaler Hyperwürfel ist. Ist h_n eine Kante, ist $V_n = h_n^d$. Weiter wird eine Fensterfunktion φ zur Hilfe genommen, welche den Würfel einheitlich auf den Ursprung setzt. Betrachte nun

$$\varphi((x - x_i)/h_n)$$

Liegt ein Punkt x_i im Würfel um x , ist die Fensterfunktion 1, liegt sie nicht drin, ist sie 0. k_n ist also:

$$k_n = \sum_{i=1}^n \varphi\left(\frac{x - x_i}{h_n}\right)$$

somit ist der Schätzer Mittelwert der Funktionen von x und den Proben x_i . Die Fensterfunktion dient somit als Interpolation - jede Probe trägt zur Schätzung ihren Abstand zu x bei. Wichtig: Die kompletten Daten werden berücksichtigt.

$$p_n(x) = \frac{1}{n} \cdot \left(\sum_{i=1}^n \frac{1}{V_n} \varphi\left(\frac{x - x_i}{h_n}\right) \right)$$

- bei der k_n nearest neighbor estimation wird k_n von n eingeschränkt, zum Beispiel durch \sqrt{n} . Hier vergrößert man den Würfel um x solange, bis die Vorlage durch n erreicht wird. Das sind dann die k_n nächsten Nachbarn von x . Wenn nun um x viele Punkte liegen, ist die Dichte im Würfel hoch und die Zelle relativ klein zu den gesamten Daten. Ist die Dichte der Punkte im Würfel gering, wächst natürlich der Würfel ein wenig mehr, erreicht aber eine gewisse Obergrenze. Das heißt nun:

$$\frac{k_n}{n} \xrightarrow[n \rightarrow \infty]{} 0$$

Wichtig, sodass $p_n(x)$ gegen $p(x)$ konvergiert.

3. In der Realität sind parametrisierte Methoden manchmal eher unpassend, da die Rohdaten in einigen Fällen die Beschaffenheit für eine unimodale Dichtefunktion nicht hergeben. Das liegt an dem einen lokalen Maximum, welches unimodale Funktionen besitzen. Manche Daten benötigen eher multimodale Dichtefunktionen, sodass sie akkurat sind. Dafür gibt es nicht-parametrische Prozeduren, welche mit den vorliegenden Verteilungen arbeitet anstatt eine Annahme über die Dichte vorzusetzen. Die Methoden schätzen eine passende Dichtefunktion und ist sie gut genug, wird sie für den Klassifizier verwendet.