

Maschinelles Lernen Blatt 5

Nikolas Zeitler, Joshua Hartmann, Alexander Diegel

June 24, 2016

1 Hidden Markov Model

2 Fragen zur Vorlesung

1. Warum gilt für $n \rightarrow \infty$ mit $k \rightarrow \inf$ und $V_n \rightarrow 0$ für den Schätzer $p_n(x)$ die Identität?

2. Was ist die Strategie der Dichteschätzung bei dem Parzenfenster-Verfahren und bei dem Nächster- Nachbar-Verfahren?

Parzenfenster-Verfahren: Man wählt eine Region V und zählt dann, wie viele Datenpunkte in dieser Region liegen.

Nächster-Nachbar-Verfahren: Man legt eine Anzahl von k Nachbarn fest, die in einem Gebiet enthalten sein müssen. Man vergrößert das Gebiet um einen Datenpunkt so lange, bis sich eben k Nachbarn in diesem Gebiet befinden.

Die Dichte ergibt sich jeweils aus dem Verhältnis aus der Anzahl k der Datenpunkte zur Größe der Region V .

3. Inwiefern handelt es sich bei beiden in der letzten Aufgabenstellung genannten Verfahren um sogenannte nicht-parametrische Methoden? Sie werden gelegentlich auch Prototypen-basierte Verfahren genannt. Erkläre, weswegen dies sinnvoll ist.

Wir bestimmen dabei keine Parameter, die die entsprechende Dichtefunktion beschreiben würden, sondern modellieren anhand von gegebenen Datenpunkten die Dichtefunktion.

Das ist deshalb sinnvoll, da man damit auch Dichtefunktionen modellieren kann, die nicht durch einfache Parameter beschreibbar ist beziehungsweise von der wir gar nicht wissen, von was für einer grundsätzlichen Gestalt sie ist.

3 Parzenfenster