

# ALG5I - Übung 01

Alen Kocaj

24. Oktober 2017

## 1 Bayesian Decision Theory

a.) Die Bayesian Decision Theory repräsentiert ein statistisches Entscheidungsverfahren basierend auf Wahrscheinlichkeiten. Die Theorie besagt, dass unter der Annahme von relevanten charakteristischen Wahrscheinlichkeiten, das Entscheidungsproblem optimal und mit minimaler Fehlerquote gelöst werden kann. Daher kann sie auch als Basis für einen Klassifizierer genommen werden. Die Theorie basiert auf dem Theorem von Bayes zur Berechnung von bedingten Wahrscheinlichkeiten wie z.B.  $P(\omega|x)$  - wie wahrscheinlich ist das Eintreten oder das Sein von  $\omega$  gegeben  $x$ .

Die zwei wichtigen Wahrscheinlichkeiten, auf welchen die Entscheidungstheorie nach Bayes aufbaut, sind die *a priori probability*  $P(\omega)$  und die *class-conditional probability density function*  $p(x|\omega)$ . Dabei wird  $\omega$  als *state of nature* - als eine Klasse oder Kategorie bezeichnet, zu der man klassifizieren kann. Würde man das Wetter klassifizieren wollen, wären z.B.  $\omega_1$  = sonnig,  $\omega_2$  = bewölkt und  $\omega_3$  = regnerisch. Die kontinuierliche Zufallsvariable  $x$  bezeichnet ein Feature, also eine Ausprägung eines Datensets, welches für die Klassifizierung verwendet werden kann. Bei dem Wetterbeispiel von vorhin wäre  $x$  z.B. die Luftfeuchtigkeit.

Die *a priori probability*  $P(\omega)$  bezeichnet das Wissen, welches wir im Vorhinein über das Eintreten einer bestimmten Kategorie oder Klasse  $\omega$ . In unserem Beispiel zum Wetter könnte man z.B. aus den Daten entnehmen, dass  $P(\omega_1) = 0.7$ , also die Wahrscheinlichkeit, dass der Tag sonnig ist. Das heißt für zukünftige Klassifizierungen würden wir uns zum Einen auf diesen Wahrscheinlichkeitswert stützen. Übrigens, es ist genauso legitim, die *a priori probability* zu schätzen bzw. anzunehmen. Es gibt Fälle, bei denen wir uns theoretisch nur auf die *a priori probability* verlassen könnten, ohne andere Wahrscheinlichkeiten zu betrachten. Dies passiert, wenn wir, gegeben den Daten, die Klassifizierung nur mit unserem Vorwissen vornehmen können. D.h. wir würden ohne Betrachtung von Features oder Eigenschaften den Tag einstufen. Logischerweise, würden wir uns für die höhere Wahrscheinlichkeit entscheiden.  $P(\omega_1 = \text{sonnig}) = 0.7 > P(\omega_2 = \text{bewölkt}) = 0.2 > P(\omega_3 = \text{regnerisch}) = 0.1$ . Natürlichweise fällt die Kurzsichtigkeit dieser Method sofort auf. Wir würden jedes Mal den Tag als sonnig klassifizieren, selbst den zu betrachtende Datensatz ein regnerischer Tag wäre.

Möchte man nun Features in die Entscheidungshilfe miteinfließen lassen, so betrachtet man alle Samples dieser Feature basierend auf den zu klassifizierenden Kategorien. Dies ergibt eine Verteilung der Samples je pro Feature pro Klasse. Daher wird diese Wahrscheinlichkeit auch als *class-conditional probability density*  $p(x|\omega)$  bezeichnet. Dieser Ausdruck lässt sich übersetzen mit der Aussage: “Gegeben eine Kategorie  $\omega$ , was ist die Verteilung der Zufallsvariable  $x$ , wobei  $x$  ein Feature repräsentiert?”. Die Verteilung der Zufallsvariable wird entweder aus den Daten entnommen oder, wie oft in der Praxis geschätzt. Der Einfachheit halber betrachten wir uns nur ein Feature. Aus unserem Beispiel von vorhin wäre dies die Luftfeuchtigkeit.