WSI für Informatik an der Karls-Eberhardt Universität Tübingen

Machine Learning

Übungsblatt 1

Benjamin Çoban 3526251 Lea Bey 3907702

11. Mai 2017

1

Aufgabe 1.1.

a) Die <u>Regression</u> sagt einen reellen Zahlenwert voraus, wogegen die <u>Klassifikation</u> eine Klasse vorhersagt.

b) Regression, da es um die Vorhersage von absoluten Zahlenwerten geht.

c) Klassifikation, da es um die Vorhersage von Geschlechterklassen geht.

d) Supervised Learning setzt ein Labeling der Beispieldaten voraus, aus denen der Algorithmus lernen kann.
Bei Unsupervised Learning sind die Beispieldaten nicht gelabelt und der Algorithmus versucht die Daten durch Clustern in verschiedene Klassen zu unterteilen.

e) i) Eine diskrete Zufallsvariable hat eine diskrete Anzhal an Werten, die sie annehmen kann.
Eine kontinuierliche Zufallsvariable kann unendlich viele Werte (of aber nur in einem beschränkten Bereich) annehmen.
Ein Ereignis ist das Ergebnis eines Zufallsexperiments.

ii) Beides sind Funktionen, die die Wahrscheinlichkeit eines Ereignisses (diskret) oder eines x-Achsen Abschnittes (kontinuierlich) abbilden.
Die Wahrscheinlichkeit P(x∈A) (A entweder diskreter Wert oder kontinuierliches Intervall) lässt sich dann mithilfe der Funktion/Dichte bestimmen.

iii) Der Erwartungswert einer Zufallsvariablen beschreibt die Zahl, die Zufallsvariable im Mittel annimmt.
Mathematisch: mit x_i Wert und p_i die zugehörige Wahrscheinlichkeit, f(x) ist die Wahrscheinlichkeitsdichte
Diskret: E(X)= $\sum_{i \in I} x_i p_i$
Kontinuierlich: E(X)=$\int_{-\infty}^{\infty} x * f(x) dx$

iv) Die <u>Varianz</u> Var(X) beschreibt die erwartete quadratische Abweichung der Zufallsvariablen von ihrem Erwartungswert.

$$\begin{aligned} Var(X) &= \sum_{x \in A} (x-\mu)^2 P([X=x]) \text{ (diskret)} \\ &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x) \mathrm{d}x \text{ (stetig)} \end{aligned}$$

Weiter gilt für die Varianz:

$$Var(X) = E((X - E(X))^2) = E(X^2) - E(X)^2 \text{ mit } Y = (X - E(X))^2$$

Also ist Y die quadratische Abweichung von X zu ihrem eigenen Erwartungswert.
Die Standardabweichung σ ist eine reelle Zahl und ist definiert als:

$$\sigma_X = \sqrt{Var(X)}$$

Sie beschreibt die Streuung einer Zufallsvariable X.

v) Die Covarianz $Cov(X, Y)$ beschreibt den monotonen Zusammenhang zweier Variablen und ist definiert als:

$$Cov(X, Y) = E[(X - E(X)) \cdot (Y - E(Y))]$$

Es gibt mehrere Ergebnisfälle:

$Cov(X, Y)$	desc
> 0	Es besteht ein monotoner Zusammenhang
$= 0$	Es besteht kein monotoner Zusammenhang
< 0	Es besteht ein gegenläufig monotoner Zusammenhang

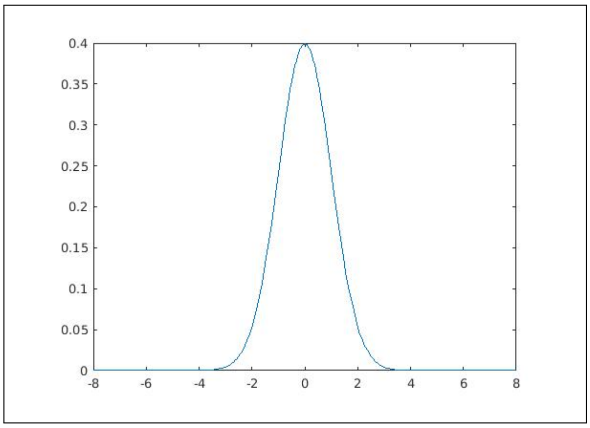
f) Für die Dichtefunktion muss gelten:

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$$
$$\Rightarrow \int_{-\infty}^{\infty} [f(x) \cdot g(x)] \neq \int_{-\infty}^{\infty} f(x)\mathrm{d}x \cdot \int_{-\infty}^{\infty} g(x)\mathrm{d}x = 1 \cdot 1 = 1$$

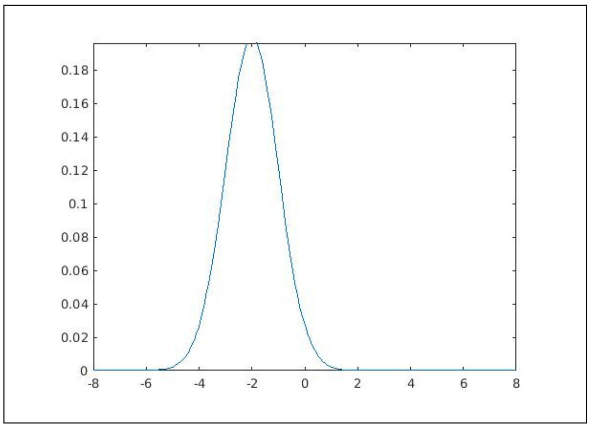
Somit entsteht bei Multiplikation keine neue Dichtefunktion.

Aufgabe 1.2.

a)



b)



Der Wert μ gibt die Verschiebung des Maxima in x-Richtung an, der Graph wird mit -2 also nach links verschoben.
Der Wert σ gibt die Stauchung bzw Streckung des Graphen an. Der Graph wird also um den Faktor 2 gestaucht.

c) Um die Wahrscheinlichkeit eines Events innerhalb eines Intervalls zu berechnen wird der Flächeninhalt der Dichtefunktion in diesem Intervall berechnet.
Also: $P(X \in [a, b]) = \int_b^a f(x)\mathrm{d}x$

d) Im Intervall $[\mu \pm \sigma]$ sind 68,27% aller Werte.
Im Intervall $[\mu \pm 2\sigma]$ sind 95,45% aller Werte.
Im Intervall $[\mu \pm 3\sigma]$ sind 99,73% aller Werte.

e) Die Wahrscheinlichkeitsdichtefunktion kann auch Werte größer als 1 annehmen, solange sie danach normiert ist, solange also das Integral von - Unendlich bis Unendlich gleich 1 ist (per Definition).