

Übungsblatt

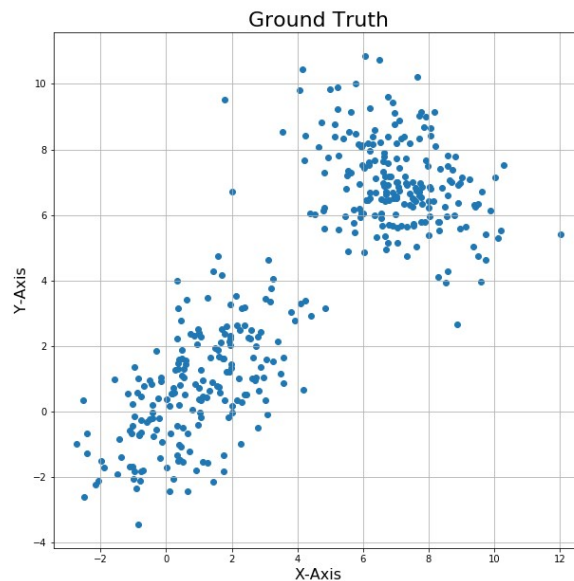
Grundlagen der Künstlichen Intelligenz

24.11.2023, DHBW Lörrach

- Clustering-Verfahren -

Expectation-Maximization-Algorithmus

Nehmen Sie an, Kunden entscheiden sich für ($y = 1$) oder gegen ($y = 0$) den Kauf eines Produkts basierend auf den beiden Features x (X-Achse) und y (Y-Achse):



Die Punkte stellen die Kundenmenge d dar.

1. Erläutern Sie die Erzeugung der Parameter $m1$, $m2$, $cov1$ und $cov2$ in den Code-Zeilen 1-4 im Initialisierungsschritt des Expectation-Maximization Algorithmus.

```
1 m1 = random.choice(d)
2 m2 = random.choice(d)
3 cov1 = np.cov(np.transpose(d))
4 cov2 = np.cov(np.transpose(d))
5 pi = 0.5 #Apriori Wahrscheinlichkeit für bzw. gegen den Kauf des Produkts
```

- m_i : Erwartungswert des i -ten Gauss-Modus, der randomisiert initialisiert wird
- cov_i : Kovarianzmatrix des i -ten Gauss-Modus, die basierend auf der gesamten Punkteverteilung d initialisiert wird

2. Geben Sie ein Beispiel für die beiden Features x und y sowie deren Parameter m_1 und m_2 .

- Beispiel: Der Erwartungswert $m_1 = 100$ Euro gibt an, bei welchem Preis die Normalverteilungskurve ihr Maximum annimmt. Analog: $m_2 = 150$ cm als Konfektionsgröße für Bekleidung.

3. Erläutern Sie den Erwartungsschritt im EM-Algorithmus und gehen dabei auf die Erzeugung der Wahrscheinlichkeitsdichtefunktion (*pdf*) in den Code-Zeilen 8 und 9 ein.

```
1  def Estep(lis1):
2      m1=lis1[0]
3      m2=lis1[1]
4      cov1=lis1[2]
5      cov2=lis1[3]
6      pi=lis1[4]
7
8      pt2 = multivariate_normal.pdf(d, mean=m2, cov=cov2)
9      pt1 = multivariate_normal.pdf(d, mean=m1, cov=cov1)
10     w1 = pi * pt2
11     w2 = (1-pi) * pt1
12     eval1 = w1/(w1+w2)
13
14     return(eval1)
```

- Erwartungsschritt: Zuordnung der Datenpunkte zum entsprechenden Gaussmodus
- Wahrscheinlichkeitsdichtefunktion in den Code-Zeilen 8 und 9: Die Daten werden unter der Annahme einer Normalverteilung modelliert und anschließend gewichtet und einem Gauss-Modus zugeordnet.