Grundlagen der künstlichen Intelligenz: Hausaufgabe 7

Tom Nick - 340528
Niklas Gebauer - 340942
Leonard Witte - 341457
Johannes Herrmann - 341091

Aufgabe 1

a.)

$$
\begin{aligned}
&P(''TCGCGA'' \mid X = f) \\
&= P(Y_1 = T \mid X = f) \cdot P(Y_2 = C \mid Y_1 = T, X = f) \cdot P(Y_3 = G \mid Y_2 = C, X = f) \\
&\cdot P(Y_4 = C \mid Y_3 = G, X = f) \cdot P(Y_5 = G \mid Y_4 = C, X = f) \cdot P(Y_6 = A \mid Y_5 = G, X = f) \\
&= 0,25 \cdot 0,24 \cdot 0,08 \cdot 0,25 \cdot 0,08 \cdot 0,25 \\
&= \frac{3}{125000} \approx 0,000024
\end{aligned}
$$

$$
\begin{aligned}
&P(''TCGCGA'' \mid X = w) \\
&= P(Y_1 = T \mid X = w) \cdot P(Y_2 = C \mid Y_1 = T, X = w) \cdot P(Y_3 = G \mid Y_2 = C, X = w) \\
&\cdot P(Y_4 = C \mid Y_3 = G, X = w) \cdot P(Y_5 = G \mid Y_4 = C, X = w) \cdot P(Y_6 = A \mid Y_5 = G, X = w) \\
&= 0,25 \cdot 0,36 \cdot 0,27 \cdot 0,34 \cdot 0,27 \cdot 0,16 \\
&\approx 0,000357
\end{aligned}
$$

Gemäß der Maximum-Likelihood-Methode würden wir uns also für das Modell $X = w$ entscheiden, also sagen, dass die DNA-Sequenz teil einer CpG-Insel ist.

b.) Wir müssen berechnen:

$$P(X = w \mid ''TCGCGA'') = \frac{P(''TCGCGA'' \mid X = w) \cdot P(X = w)}{\sum_X P(''TCGCGA'' \mid X) * P(X)}$$

Dazu berechnen wir vorerst:

$$
\begin{aligned}
&P(''TCGCGA'' \mid X = f) * P(X = f) \\
&= \frac{3}{125000} \cdot 0,8 = \frac{3}{156250} \approx 0,0000192
\end{aligned}
$$

$$
\begin{aligned}
&P(''TCGCGA'' \mid X = w) * P(X = w) \\
&= 0,000357 \cdot 0,2 \approx 0,0000714
\end{aligned}
$$

Nun können wir die Posterior-Wahrscheinlichkeit berechnen:

$$P(X = w \mid ''TCGCGA'') = \frac{0,0000714}{\frac{3}{156250} + 0,0000714} = \frac{119}{151} \approx 0,7881$$

Die Posterior-Wahrscheinlichkeit, dass die DNA-Sequenz zu einer CpG-Insel gehört, ist also ungefähr 78, 81%.

c.)

$$
\begin{aligned}
P(Y_7 = g \mid ''TCGCGA'') &= \sum_X P(Y_7 = g \mid Y_6 = a, X) \cdot P(X \mid ''TCGCGA'') \\
&= 0,29 \cdot (1 - \frac{119}{151}) + 0,43 \cdot \frac{119}{151} = \frac{1209}{3020} \approx 0,40
\end{aligned}
$$

d.) Ja, die Vorhersage würde sich ändern: $X = w$ ist laut MAP-Methode das wahrscheinlichste Modell. Damit wäre die Vorhersage

$$P(Y_7 = g \mid Y_6 = a, X = w) = 0,43$$

etwas optimistischer, dass als nächstes Nukleotid g auftritt, als unsere Vorhersage aus c), die beide Modelle berücksichtigt.

Aufgabe 2

a.) Es handelt sich um unabhängige Zufallsexperimente mit einer festen Folge, also kann für die Likelihood eine Binomialverteilung ohne den Parameter $\binom{n}{k}$ angenommen werden:

$$P(\text{k Erfolge, 3 Fehlschläge}) = p^k \cdot (1-p)^3$$

b.) Sei X der erwartete Gewinn pro Runde. Da der Einsatz 1 Euro und die Auszahlung 2 Euro sind, ist der Gewinn beim Ereignis Kopf -1 Euro (also ein Verlust von einem Euro) und beim Ereignis Zahl 1 Euro. Somit ergibt sich der erwartete Gewinn:

$$E(X) = 1 \cdot 0,7 + (-1) \cdot 0,3 = 0,4$$

Pro Runde kann also ein Gewinn von 40 Cent erwartet werden.

c.) Gemäß der Übung wissen wir für die Binomialverteilung (ohne Parameter $\binom{n}{k}$) die Form der beiden Hypothesen. Es gilt:
Maximum-Likelihood-Hypothese maximal mit:

$$p = \frac{k}{n}$$

Maximum-a-posteriori-Hypothese maximal mit:

$$p = \frac{k + (\alpha - 1)}{n + (\alpha - 1) + (\beta - 1)}$$

Wenn wir nun also die Hyperparameter $\alpha = \beta = 1$ setzen, stimmen beide Hypothesen überein:

$$\frac{k + 1 - 1}{n + 1 - 1 + 1 - 1} = \frac{k}{n}$$

d.) Für den Posterior des Wurfkunststücks gilt:

$$\begin{aligned} P(p \mid k, \alpha, \beta) &= \text{Likelihood} \cdot \text{Prior} \\ &= p^k \cdot (1-p)^3 \cdot B(\alpha, \beta) \cdot p^{\alpha-1} \cdot (1-p)^{\beta-1} \\ &= \gamma \cdot p^{k+\alpha-1} \cdot (1-p)^{\beta+2} \\ &\text{mit } \gamma = B(\alpha, \beta) \end{aligned}$$

Wir nehmen den Logarithmus:

$$logP(p \mid k, \alpha, \beta) = (k + \alpha - 1) \cdot logp + (\beta + 2) \cdot log(1-p) + log\gamma$$

Es ergibt sich nun für die Maximum-a-posteriori-Hypothese:

$$\begin{aligned} \frac{\partial logP(p \mid k, \alpha, \beta)}{\partial p} &= 0 \\ \Leftrightarrow \frac{k + \alpha - 1}{p} - \frac{\beta + 2}{1 - p} &= 0 \\ \Leftrightarrow \frac{p}{k + \alpha - 1} &= \frac{1}{\beta + 2} - \frac{p}{\beta + 2} \\ \Leftrightarrow p \cdot \left(\frac{1}{k + \alpha - 1} + \frac{1}{\beta + 2}\right) &= \frac{1}{\beta + 2} \\ \Leftrightarrow P &= \frac{1}{\beta + 2 \cdot \left(\frac{1}{k+\alpha-1} + \frac{1}{\beta+2}\right)} \\ \Leftrightarrow p &= \frac{1}{\frac{\beta+2+k+\alpha-1}{k+\alpha-1}} \\ \Leftrightarrow p &= \frac{k + \alpha - 1}{k + \alpha + \beta + 1} \end{aligned}$$

e.) Mit der Gleichung aus d) ergibt sich:

$$p = \frac{7 + 3 - 1}{7 + 3 + 4 + 1} = \frac{9}{15} = 0,6$$