

Aufgabe HMM_2:

Nach der im Abschnitt 16.5 beschriebenen Vorgehensweise folgt aus dem multiplen Sequenzalignment (MSA) die folgende Zuordnung von Zuständen:

0	A	C	Y	L	V	M	V	N	C
1	A	C	Y	I	I	-	V	-	C
2	G	C	-	I	V	-	I	N	C
3	G	C	-	L	V	-	V	-	C
4	A	C	Y	L	L	M	I	A	C
5	G	C	Y	L	-	-	I	A	C
	M_1	M_2	M_3 D_3	M_4	M_5 D_5	I_5	M_6	M_7 D_7	M_8

Es ergeben sich demnach 8 Match- und ebenso viele Insertions- und Deletions-Zustände. Aus der Zuordnung bestimmen wir für jede Zeile des MSAs einen Pfad:

```

A[0][.]  A  C  Y  L  V  M  V  N  C
          M1 M2 M3 M4 M5 I5 M6 M7 M8

A[1][.]  A  C  Y  I  I  -  V  -  C
          M1 M2 M3 M4 M5      M6 D7 M8

A[2][.]  G  C  -  I  V  -  I  N  C
          M1 M2 D3 M4 M5      M6 M7 M8

A[3][.]  G  C  -  L  V  -  V  -  C
          M1 M2 D3 M4 M5      M6 D7 M8

A[4][.]  A  C  Y  L  L  M  I  A  C
          M1 M2 M3 M4 M5 I5 M6 M7 M8

A[5][.]  G  C  Y  L  -  -  I  A  C
          M1 M2 M3 M4 D5      M6 M7 M8

```

Nun können die Übergangs- und Emissionswahrscheinlichkeiten unter Verwendung der Laplaceschen Regel geschätzt werden:

Aus dem Zustand M_2 ist der Übergang nach M_3 , D_3 und I_2 möglich. Aufgrund der Laplaceschen Regel werden sämtliche Werte mit 1 vorbesetzt. Es folgt dann für die derart korrigierte Anzahl von Übergängen: $M_2 M_3 = 5$, $M_2 I_2 = 1$, $M_2 D_3 = 3$ und daraus folgen die Übergangswahrscheinlichkeiten: $a_{M_2 M_3} = 5 / 9$, $a_{M_2 I_2} = 1 / 9$, $a_{M_2 D_3} = 3 / 9$.

Die Emissionswahrscheinlichkeiten schätzen wir unter Verwendung der Laplaceschen Regel für den Match-Zustand M_2 zu $e_{M_2}(C) = 7 / 26$ und $e_{M_2}(\sigma) = 1 / 26$ für alle anderen 19 Aminosäuren, sowie für den Insertionszustand I_2 : $e_{I_2}(\sigma) = 1 / 20$ für alle 20 Aminosäuren.

Analog gilt für M_5 : $a_{M_5 M_6} = 4 / 8$, $a_{M_5 I_5} = 3 / 8$, $a_{M_5 D_6} = 1 / 8$ und die Emissionswahrscheinlichkeiten schätzen wir wie folgt für den Match-Zustand M_5 :

$e_{M_5}(V) = 4 / 25$, $e_{M_5}(I) = 2 / 25$, $e_{M_5}(L) = 2 / 25$, $e_{M_5}(\sigma) = 1 / 25$ für alle anderen 17 Aminosäuren und den Insertionszustand I_5 : $e_{I_5}(M) = 3 / 22$, $e_{I_5}(\sigma) = 1 / 22$ für alle anderen 19 Aminosäuren.

Hinweis: Natürlich gibt es andere, dem biologischen Problem besser angepasste Alternativen, die Wahrscheinlichkeiten zu korrigieren. Im Basistext wird der MAP-Algorithmus vorgestellt und erläutert, wie Häufigkeiten mit Dirichlet-Dichten geschätzt werden können. Hier geht es uns jedoch darum, das (einfachste) Prinzip zu verstehen.