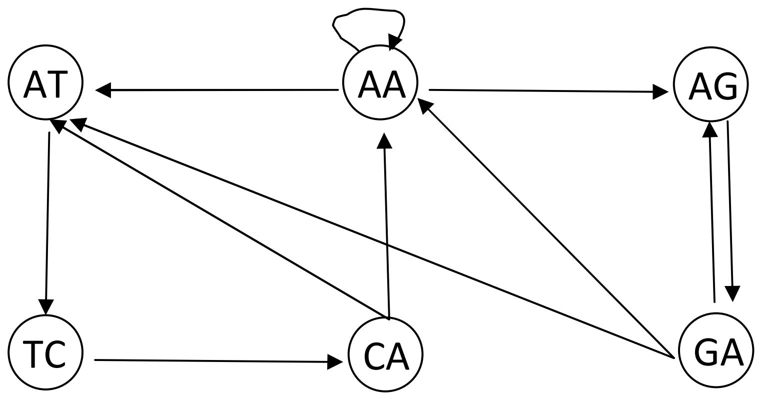
14. a)



b) AGATAG, since TA is not in the state space.

15. Let $S = \{s_1, \dots, s_N\}$ denote the state space, and $X_1^T = X_1, \dots, X_T$ and Y_1^T the hidden state sequence and the observed sequence respectively.

a) *Forward*: the probability of the observed sequence $P(Y_1^T)$

Viterbi: the probability of the most likely state sequence

$\max_{X_1^T} P(X_1^T | Y_1^T) = \max_{X_1^T} P(X_1^T, Y_1^T)$

b) Let π_i, a_{ij}, and $b_i(Y_t | Y_1^{t-1})$ denote the initial probabilities, transition probabilities, and emission probabilities respectively.

Forward:

initiation $\alpha_i(0) = \pi_i$

recursion $\alpha_i(t) = P(Y_1^t, X_t = i) = \sum_{j \in S} a_{ji} b_i(Y_t | Y_1^{t-1}) \alpha_j(t-1)$

termination $\alpha_i(T+1) = \sum_{j \in S} a_{ji} \alpha_j(T)$

Viterbi:

initiation $\delta_i(0) = \pi_i$

recursion $\delta_i(t) = \max_{1 \le j \le N} P(Y_1^T, X_1^{t-1}, X_t = i) = \max_{1 \le j \le N} a_{ji} b_i(Y_t | Y_1^{t-1}) \delta_j(t-1)$

termination $\delta_i(T+1) = \max_{1 \le j \le N} a_{ji} \delta_j(T)$

Store pointers for each $\delta_i(t)$ to previous state

$\psi_i(t) = \text{argmax}_{1 \le j \le N} a_{ji} b_i(Y_t | Y_1^{t-1}) \delta_j(t-1)$

c) Forward: no traceback is needed. Probability is calculated upon termination

$P(Y_1^T) = \sum_j \alpha_j(T+1)$

Viterbi: follow the stored pointers ψ, starting in

$X_{T+1}^* = \text{argmax}_{1 \le j \le N} \delta_j(T+1)$

backtrack using

$X_t^* = \psi_{X_{t+1}^*}(t+1)$, $t = T, T-1, \dots, 0$

16. a) For an observed sequence $Y_1^T = Y_1, \dots, Y_T$ the log-odds ratio is given by

$$L(Y_1^T) = \log \frac{P(Y_1^T|S_+)}{P(Y_1^T|S_-)} \begin{cases} > \eta & \text{then } S_+ \\ < \eta & \text{then } S_- \end{cases}$$

for some threshold η.

b) (Dessvärre har jag lärt ut lite fel på kursen här, vilket ett par studenter har uppmärksammat, så eventuellt kommer två lite olika lösningar presenteras.)

Utlärd version:

$\delta_i(0) = \pi_i$, $i = 1, \dots, N$ (detta är fel)

$\delta_i(t) = \max_{1 \le j \le N} a_{ji} b_i(Y_t) \delta_j(t-1)$

$\delta_i(T+1) = \max_{1 \le j \le N} a_{ji} \delta_j(T)$

Korrekt version:

$\delta_i(1) = \pi_i b_i(Y_1)$

$\delta_i(t)$ som ovan för $t = 2, \dots, T$, och samma termination

Beräkning enligt den utlärda (inkorrekta) versionen blir:

$\pi = \{\pi_{S_+}, \pi_{S_-}\} = \{0.5, 0.5\}$

$a_{ij} = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}$ för (S_+, S_-)

$b_i(Y_t) = \begin{pmatrix} 0.4 & 0.4 & 0.1 & 0.1 \\ 0.3 & 0.3 & 0.2 & 0.2 \end{pmatrix}$ för rader S_+, S_-, och kolumner (A, C, G, T)

$Y_1^4 = GTCT$

$\delta_{S_+}(0) = \delta_{S_-}(0) = 0.5$

$Y_1 = G$:

$\delta_{S_+}(1) = \max\{0.9 \cdot 0.1 \cdot 0.5, 0.1 \cdot 0.1 \cdot 0.5\} = \max\{0.045, 0.005\} = 0.045$,

$\psi_{S_+}(1) = S_+$

$\delta_{S_-}(1) = \max\{0.1 \cdot 0.2 \cdot 0.5, 0.9 \cdot 0.2 \cdot 0.5\} = \max\{0.01, 0.09\} = 0.09$

$\psi_{S_-}(1) = S_-$

$Y_2 = T$:

$\delta_{S_+}(2) = \max\{0.9 \cdot 0.1 \cdot 0.045, 0.1 \cdot 0.1 \cdot 0.09\} = \max\{0.0405, 0.005\} = 0.0405$,

$\psi_{S_+}(2) = S_+$

$\delta_{S_-}(2) = \max\{0.1 \cdot 0.2 \cdot 0.045, 0.9 \cdot 0.2 \cdot 0.09\} = \max\{0.0009, 0.0162\} = 0.0162$

$\psi_{S_-}(2) = S_-$

$Y_3 = C$:

$\delta_{S_+}(3) = \max\{0.9 \cdot 0.4 \cdot 0.00405, 0.1 \cdot 0.4 \cdot 0.0162\} = \max\{0.001458, 0.000648\} = 0.0001458$

$\psi_{S_+}(3) = S_+$

$\delta_{S_-}(3) = \max\{0.1 \cdot 0.3 \cdot 0.00405, 0.9 \cdot 0.3 \cdot 0.0162\} = \max\{0.0001215, 0.004374\} =$
0.004374
$\psi_{S_-}(3) = S_-$

$Y_4 = T$:
$\delta_{S_+}(4) = \max\{0.9 \cdot 0.1 \cdot 0.001458, 0.1 \cdot 0.1 \cdot 0.004374\} =$
$\max\{0.00013122, 0.000078732\} = 0.00013122$
$\psi_{S_+}(4) = S_+$
$\delta_{S_-}(4) =$
$\max\{0.1 \cdot 0.2 \cdot 0.001458, 0.9 \cdot 0.2 \cdot 0.004374\} = \max\{0.00002916, 0.00078732\} =$
0.00078732
$\psi_{S_-}(4) = S_-$

Termination:
$\delta_{S_+}(5) =$
$\max\{0.9 \cdot 0.00013122, 0.1 \cdot 0.00078732\} = \max\{0.000118098, 0.000078732\} =$
0.000118098
$\delta_{S_-}(5) =$
$\max\{0.1 \cdot 0.00013122, 0.9 \cdot 0.00078732\} = \max\{0.000013122, 0.000708588\} =$
0.000708588

Backtrack:
$X_5^* = S_-$
Optimal state path: $S_-S_-S_-S_-$