1 Problem 1

1.1

$$
\begin{aligned}
\mathcal{Q}(\boldsymbol{\theta};\boldsymbol{\theta}^{(t)}) &= \sum_{n=1}^{N} \mathbb{E}_{z_n \sim q_n^{(t)}}[\ln P(\mathbf{x}_n, z_n; \boldsymbol{\theta})] \\
&= \dots \\
&= \sum_{n=1}^{N}\sum_{k=1}^{K} \gamma_{nk}(\ln P(z_n;\theta) + \ln P(x_n|z_n;\theta))
\end{aligned}
$$

For fixed k and b_k,

$$
\begin{aligned}
\mu_k^* &= \mathrm{argmax}_{\mu_k} \sum_{n=1}^{N} \gamma_{nk} \ln \mathcal{L}(x_n; \mu_k, b_k) \\
&= \mathrm{argmax}_{\mu_k} \sum_{n=1}^{N} \gamma_{nk} \ln \left(\frac{1}{2b_k} \exp\left(\frac{-|x_n - \mu_k|}{b_k} \right) \right) \\
&= \mathrm{argmax}_{\mu_k} \sum_{n=1}^{N} \gamma_{nk} \left[\ln\left(\frac{1}{2b_k} \right) - \frac{|x_n - \mu_k|}{b_k} \right] \\
&= \mathrm{argmax}_{\mu_k} \sum_{n=1}^{N} -\gamma_{nk}|x_n - \mu_k| \\
&= \mathrm{argmax}_{\mu_k} \sum_{j:(x_j-\mu_k)<0}^{J} \gamma_{jk}(x_j - \mu_k) - \sum_{i:(x_i-\mu_k)\geq 0}^{I} \gamma_{ik}(x_i - \mu_k) \qquad \textit{where } (I + J = N) \\
&= \mathrm{argmax}_{\mu_k} \underbrace{\sum_{j:(x_j-\mu_k)<0}^{J} \gamma_{jk}\mu_k - \sum_{i:(x_i-\mu_k)\geq 0}^{I} \gamma_{ik}\mu_k}_{:= Q}
\end{aligned}
$$

$$
\frac{dQ}{d\mu_k} = \sum_{j:(x_j-\mu_k)<0}^{J} \gamma_{jk} - \sum_{i:(x_i-\mu_k)\geq 0}^{I} \gamma_{ik} = 0 \implies \sum_{j:(x_j-\mu_k)<0}^{J} \gamma_{jk} = \sum_{i:(x_i-\mu_k)\geq 0)}^{I} \gamma_{ik}
$$

The optimum location parameter is set such that the sum of soft assignments for data points less than μ_k is equal to that of data points greater than μ_k. In other words,

$$
\boxed{\mu_k^* = \textit{median}\{\gamma_{1k}x_1, \dots, \gamma_{nk}x_n\}}
$$

1.2

For fixed k and μ_k,

$$\begin{aligned}
b_k^* &= \text{argmax}_{b_k} \sum_{n=1}^{N} \gamma_{nk} \ln \mathcal{L}(x_n; \mu_k, b_k) \\
&= \text{argmax}_{b_k} \sum_{n=1}^{N} \gamma_{nk} \ln \left(\frac{1}{2b_k} \exp \left(\frac{-|x_n - \mu_k|}{b_k} \right) \right) \\
&= \text{argmax}_{b_k} \sum_{n=1}^{N} \gamma_{nk} \left[\ln(1) - \ln(2) - \ln(b) - \frac{|x_n - \mu_k|}{b_k} \right] \\
&= \text{argmax}_{b_k} \sum_{n=1}^{N} \gamma_{nk} \left[- \ln(b) - \frac{|x_n - \mu_k|}{b_k} \right]
\end{aligned}$$

$$\begin{aligned}
\frac{d}{db_k} \left(- \ln(b) - \frac{|x_n - \mu_k|}{b_k} \right) \Bigg|_{\mu_k} &= \sum_{n=1}^{N} \gamma_{nk} \left[\frac{-1}{b_k} + \frac{|x_n - \mu_k|}{b_k^2} \right] = 0 \\
\implies b_k \sum_{n=1}^{N} \gamma_{nk} &= \sum_{n=1}^{N} \gamma_{nk} |x_n - \mu_k| \\
\implies & \boxed{b_k = \frac{\sum_{n=1}^{N} \gamma_{nk} |x_n - \mu_k|}{\sum_{n=1}^{N} \gamma_{nk}}}
\end{aligned}$$

1.3

Step 0: Initialize $\boldsymbol{\theta}^{(0)} = (\boldsymbol{\mu}, \boldsymbol{b}, \boldsymbol{\omega})$

Step 1 (E-step): Update soft assignment γ_{nk}

Step 2 (M-step): Update model parameters ω_k, μ_k, b_k

Step 3: If $P(\boldsymbol{x}, \boldsymbol{\theta})$ increases, back to step 1, else end loop

The difference between the LMM EM algo and GMM EM algo is that the former does not have a closed form solution to compute the optimal location parameter, whereas the latter does. A search algorithm may be required to compute the median.

2 Problem 2

2.1

$$
\begin{aligned}
\mathbf{x}_n^T \mathbf{u}_i &= \Big(\sum_{j=1}^{D} \alpha_{nj} \mathbf{u}_j\Big)^T \mathbf{u}_i \\
&= \sum_{j=1}^{D} \alpha_{nj} \mathbf{u}_j^T \mathbf{u}_i \\
&= \alpha_{n1} \cancelto{0}{\mathbf{u}_1^T \mathbf{u}_i} + \cdots + \alpha_{ni} \mathbf{u}_i^T \mathbf{u}_i + \cdots + \alpha_{nD} \cancelto{0}{\mathbf{u}_D^T \mathbf{u}_i} \\
&= \alpha_{ni} \mathbf{u}_i^T \mathbf{u}_i \\
&= \alpha_{ni}
\end{aligned}
$$

2.2

$$
\begin{aligned}
J &= \frac{1}{N} \sum_{n=1}^{N} \|\mathbf{x}_n - \widetilde{\mathbf{x}}_n\|^2 \\
&= \frac{1}{N} \sum_{n=1}^{N} (\mathbf{x}_n - \widetilde{\mathbf{x}}_n)^T (\mathbf{x}_n - \widetilde{\mathbf{x}}_n) \\
&= \frac{1}{N} \Big(\sum_{n=1}^{N} \mathbf{x}_n^T \mathbf{x}_n - 2 \sum_{n=1}^{N} \widetilde{\mathbf{x}}_n^T \mathbf{x}_n + \sum_{n=1}^{N} \widetilde{\mathbf{x}}_n^T \widetilde{\mathbf{x}}_n \Big) \\
\implies \operatorname{argmin}_{z_{ni}} J &= \operatorname{argmin}_{z_{ni}} \underbrace{\Big(-2 \sum_{n=1}^{N} \widetilde{\mathbf{x}}_n^T \mathbf{x}_n + \sum_{n=1}^{N} \widetilde{\mathbf{x}}_n^T \widetilde{\mathbf{x}}_n \Big)}_{:= F}
\end{aligned}
$$

For fixed n, and fixed i $\leq$ M,

$$
\begin{aligned}
\frac{\partial F}{\partial z_{ni}} &= \left(\frac{\partial F}{\partial \widetilde{\mathbf{x}}_n} \right)^T \frac{\partial \widetilde{\mathbf{x}}_n}{\partial z_{ni}} \\
&= (-2\mathbf{x}_n + 2\widetilde{\mathbf{x}}_n)^T \mathbf{u}_i \\
&= -2\mathbf{x}_n^T \mathbf{u}_i + 2\widetilde{\mathbf{x}}_n^T \mathbf{u}_i = 0 \\
\implies \mathbf{x}_n^T \mathbf{u}_i &= \widetilde{\mathbf{x}}_n^T \mathbf{u}_i \\
&= \Big(\sum_{j=1}^{M} z_{nj} \mathbf{u}_j + \sum_{k=M+1}^{D} b_k \mathbf{u}_k \Big)^T \mathbf{u}_i \\
&= \sum_{j=1}^{M} z_{nj} \mathbf{u}_j^T \mathbf{u}_i + \sum_{k=M+1}^{D} b_k \cancelto{0}{\mathbf{u}_k^T \mathbf{u}_i} \qquad (\mathbf{u}_k^T \mathbf{u}_i = 0 \ \ \forall i \neq k \in \{M+1, ..., D\}) \\
&= z_{n1} \cancelto{0}{\mathbf{u}_1^T \mathbf{u}_i} + \cdots + z_{ni} \mathbf{u}_i^T \mathbf{u}_i + \cdots + z_{nM} \cancelto{0}{\mathbf{u}_M^T \mathbf{u}_i} \\
&= z_{ni} \\
\implies \boxed{z_{ni}^* = \mathbf{x}_n^T \mathbf{u}_i}
\end{aligned}
$$

The derivation above implicitly assumes $i \neq 1$ and $i \neq M$ for illustration purposes. It is without loss of generality the solution is the same if $i = 1$ or $i = M$.

2.3

From 2.2, we had

$$\text{argmin}_{b_i} J = \text{argmin}_{b_i} \underbrace{\big(-2 \sum_{n=1}^{N} \widetilde{\mathbf{x}}_n^T \mathbf{x}_n + \sum_{n=1}^{N} \widetilde{\mathbf{x}}_n^T \widetilde{\mathbf{x}}_n \big)}_{:=\, F}$$

For fixed i $>$ M,

$$\begin{aligned}
\frac{\partial F}{\partial b_i} &= \left(\frac{\partial F}{\partial \sum_{n=1}^{N} \widetilde{\mathbf{x}}_n} \right)^T \frac{\partial \sum_{n=1}^{N} \widetilde{\mathbf{x}}_n}{\partial b_i} \\
&= \big(\sum_{n=1}^{N} (-2\mathbf{x}_n + 2\widetilde{\mathbf{x}}_n) \big)^T \mathbf{u}_i \\
&= -2 \sum_{n=1}^{N} \mathbf{x}_n^T \mathbf{u}_i + 2 \sum_{n=1}^{N} \widetilde{\mathbf{x}}_n^T \mathbf{u}_i = 0
\end{aligned}$$

$$\begin{aligned}
\implies \sum_{n=1}^{N} \mathbf{x}_n^T \mathbf{u}_i &= \sum_{n=1}^{N} \widetilde{\mathbf{x}}_n^T \mathbf{u}_i \\
&= \sum_{n=1}^{N} \big(\sum_{j=1}^{M} z_{nj} \mathbf{u}_j + \sum_{k=M+1}^{D} b_k \mathbf{u}_k \big)^T \mathbf{u}_i \\
&= \sum_{n=1}^{N} \big(\sum_{j=1}^{M} z_{nj} \cancelto{0}{\mathbf{u}_j^T \mathbf{u}_i} + \sum_{k=M+1}^{D} b_k \mathbf{u}_k^T \mathbf{u}_i \big) \qquad (\mathbf{u}_j^T \mathbf{u}_i = 0 \;\; \forall i \neq j \in \{1, ..., M\}) \\
&= \sum_{n=1}^{N} \big(b_{M+1} \cancelto{0}{\mathbf{u}_{M+1}^T \mathbf{u}_i} + \cdots + b_i \mathbf{u}_i^T \mathbf{u}_i + \cdots + b_D \cancelto{0}{\mathbf{u}_D^T \mathbf{u}_i} \big) \\
&= \sum_{n=1}^{N} b_i = N b_i \\
N b_i &= \sum_{n=1}^{N} \mathbf{x}_n^T \mathbf{u}_i = \mathbf{u}_i^T \sum_{n=1}^{N} \mathbf{x}_n \\
\implies & \boxed{b_i^* = \mathbf{u}_i^T \left(\frac{1}{N} \sum_{n=1}^{N} \mathbf{x}_n \right)}
\end{aligned}$$

The derivation above implicitly assumes $i \neq M+1$ and $i \neq D$ for illustration purposes. It is without loss of generality the solution is the same if $i = M+1$ or $i = D$.

2.4

$$
\begin{aligned}
J &\triangleq \frac{1}{N}\sum_{n=1}^{N}\|\mathbf{x}_n - \widetilde{\mathbf{x}}_n\|^2 \\
&= \frac{1}{N}\sum_{n=1}^{N}\Big\|\sum_{k=1}^{D}(\mathbf{x}_n^T\mathbf{u}_k)\mathbf{u}_k - \big(\sum_{i=1}^{M} z_{ni}\mathbf{u}_i + \sum_{j=M+1}^{D} b_j\mathbf{u}_j\big)\Big\|^2 \\
&= \frac{1}{N}\sum_{n=1}^{N}\Big\|\cancel{\sum_{i=1}^{M}(\mathbf{x}_n^T\mathbf{u}_i)\mathbf{u}_i} + \sum_{j=M+1}^{D}(\mathbf{x}_n^T\mathbf{u}_j)\mathbf{u}_j - \Big(\cancel{\sum_{i=1}^{M}(\mathbf{x}_n^T\mathbf{u}_i)\mathbf{u}_i} + \sum_{j=M+1}^{D}\mathbf{u}_j^T\Big(\frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n\Big)\mathbf{u}_j\Big)\Big\|^2 \\
&= \frac{1}{N}\sum_{n=1}^{N}\Big\|\sum_{j=M+1}^{D}(\mathbf{x}_n^T\mathbf{u}_j)\mathbf{u}_j - \sum_{j=M+1}^{D}\mathbf{u}_j^T\bar{\mathbf{x}}\mathbf{u}_j\Big\|^2 \\
&= \frac{1}{N}\sum_{n=1}^{N}\Big\|\sum_{j=M+1}^{D}\big((\mathbf{u}_j^T\mathbf{x}_n)\mathbf{u}_j - \mathbf{u}_j^T\bar{\mathbf{x}}\mathbf{u}_j\big)\Big\|^2 \\
&= \frac{1}{N}\sum_{n=1}^{N}\Big\|\sum_{j=M+1}^{D}\underbrace{(\mathbf{u}_j^T(\mathbf{x}_n - \bar{\mathbf{x}}))}_{1\times 1\ \text{treat as scalar}}\mathbf{u}_j\Big\|^2 \\
&= \frac{1}{N}\sum_{n=1}^{N}\Big(\sum_{j=M+1}^{D}(\mathbf{u}_j^T(\mathbf{x}_n - \bar{\mathbf{x}}))\mathbf{u}_j\Big)^T\Big(\sum_{i=M+1}^{D}(\mathbf{u}_i^T(\mathbf{x}_n - \bar{\mathbf{x}}))\mathbf{u}_i\Big) \\
&= \frac{1}{N}\sum_{n=1}^{N}\sum_{j=M+1}^{D}(\mathbf{u}_j^T(\mathbf{x}_n - \bar{\mathbf{x}})\mathbf{u}_j^T)\sum_{i=M+1}^{D}((\mathbf{x}_n - \bar{\mathbf{x}})^T\mathbf{u}_i)\mathbf{u}_i) \\
&= \frac{1}{N}\sum_{n=1}^{N}\sum_{j=M+1}^{D}\sum_{i=M+1}^{D}\big(\mathbf{u}_j^T(\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^T\mathbf{u}_i\big)\mathbf{u}_j^T\mathbf{u}_i \\
&= \frac{1}{N}\sum_{n=1}^{N}\sum_{i=M+1}^{D}\big(\mathbf{u}_i^T(\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^T\mathbf{u}_i\big) \qquad (\mathbf{u}_j^T\mathbf{u}_i = 0\ \forall i \neq j, \mathbf{u}_j^T\mathbf{u}_i = 1\ for\ i = j) \\
&= \sum_{i=M+1}^{D}\mathbf{u}_i^T\Big[\frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \bar{\mathbf{x}})(\mathbf{x}_n - \bar{\mathbf{x}})^T\Big]\mathbf{u}_i \\
&= \sum_{i=M+1}^{D}\mathbf{u}_i^T\mathbf{S}\mathbf{u}_i
\end{aligned}
$$

2.5

$$\min_{\mathbf{u}_i} \sum_{j=M+1}^{D} \mathbf{u}_j^T \mathbf{S} \mathbf{u}_j \qquad s.t. \; \|\mathbf{u}_i\|^2 = 1$$

$$\implies \mathcal{L} = \sum_{j=M+1}^{D} \mathbf{u}_j^T \mathbf{S} \mathbf{u}_j - \lambda_i(\|\mathbf{u}_i\|^2 - 1)$$

For fixed $i : M < i \leq D$,

$$\begin{aligned}
&\frac{\partial \mathcal{L}}{\partial \mathbf{u}_i} = \mathbf{S}\mathbf{u}_i + \mathbf{S}^T\mathbf{u}_i - 2\lambda_i \mathbf{u}_i = 0 \\
\implies &2\mathbf{S}\mathbf{u}_i = 2\lambda_i \mathbf{u}_i \qquad\qquad (\mathbf{S}^T = \mathbf{S} \; \textit{because } \mathbf{S} \; \textit{symmetric}) \\
\implies &\boxed{\mathbf{S}\mathbf{u}_i^* = \lambda_i \mathbf{u}_i^*}
\end{aligned}$$

As shown above, we can express the solution for optimal $\mathbf{u}_i^*$ as an eigenvector of $\mathbf{S}$, with corresponding eigenvalue $\lambda_i, \;\; \forall i : M < i \leq D$.

3 Problem 3

3.1

$$P(X_{t+1}|O_{1:t}) = \sum_s P(X_{t+1}, X_t = s|O_{1:t})$$

3.2

For fixed state s,

$$\begin{aligned}
P(X_{t+1}, X_t = s|O_{1:t}) &= \sum_{s'} P(X_{t+1}, X_t = s, X_{t-1} = s'|O_{1:t}) \\
&= \sum_{s'} P(X_{t+1}|X_t = s, X_{t-1} = s', \cancel{O_{1:t}})P(X_t = s, X_{t-1} = s'|O_{1:t}) \\
&= \sum_{s'} P(X_{t+1}|X_t = s, X_{t-1} = s')P(X_t = s, X_{t-1} = s'|O_{1:t})
\end{aligned}$$

3.3

$$P(X_{t+1}|O_{1:t+1}) = \sum_s P(X_{t+1}, X_t = s|O_{1:t+1})$$

3.4

$$\begin{aligned}
P(X_{t+1}, X_t|O_{1:t+1}) &= P(X_{t+1}, X_t|O_{1:t}, O_{t+1}) \\
&= P(X_{t+1}, X_t, O_{t+1}|O_{1:t})/P(O_{t+1}|O_{1:t}) \\
&= P(O_{t+1}|\cancel{O_{1:t}}, X_{t+1}, \cancel{X_t})P(X_{t+1}, X_t|O_{1:t})/P(O_{t+1}|O_{1:t}) \\
&= P(O_{t+1}|X_{t+1})P(X_{t+1}, X_t|O_{1:t})/P(O_{t+1}|O_{1:t})
\end{aligned}$$

The denominator can be computed via forward messaging.

4 Problem 4

4.1

To find the probability of the observed sequence, I compute the forward messages iteratively.

$$\alpha_1(1) = \pi_1 b_{1A} = 0.6 \cdot 0.4 = 0.24$$
$$\alpha_2(1) = \pi_2 b_{2A} = 0.4 \cdot 0.2 = 0.08$$

$$\alpha_1(2) = b_{1G}(a_{11}\alpha_1(1) + a_{21}\alpha_2(1)) = 0.4(0.7 \cdot 0.24 + 0.2 \cdot 0.08) = 0.0736$$
$$\alpha_2(2) = b_{2G}(a_{12}\alpha_1(1) + a_{22}\alpha_2(1)) = 0.2(0.3 \cdot 0.24 + 0.8 \cdot 0.08) = 0.0272$$

$$\alpha_1(3) = b_{1C}(a_{11}\alpha_1(2) + a_{21}\alpha_2(2)) = 0.1(0.7 \cdot 0.0736 + 0.2 \cdot 0.0272) = 0.005696$$
$$\alpha_2(3) = b_{2C}(a_{12}\alpha_1(2) + a_{22}\alpha_2(2)) = 0.3(0.3 \cdot 0.0736 + 0.8 \cdot 0.0272) = 0.013152$$

$$\alpha_1(4) = b_{1T}(a_{11}\alpha_1(3) + a_{21}\alpha_2(3)) = 0.1(0.7 \cdot 0.005696 + 0.2 \cdot 0.013152) = 0.00066176$$
$$\alpha_2(4) = b_{2T}(a_{12}\alpha_1(3) + a_{22}\alpha_2(3)) = 0.3(0.3 \cdot 0.005696 + 0.8 \cdot 0.013152) = 0.00366912$$

$$P(O_{1:4} = [AGCT]; \theta) = \alpha_1(4) + \alpha_2(4) = \boxed{0.00433}$$

4.2

$$
\begin{aligned}
\delta_1(1) &= b_{1A}\pi_1 = 0.4 \cdot 0.6 = 0.24 \\
\delta_2(1) &= b_{2A}\pi_2 = 0.2 \cdot 0.4 = 0.08
\end{aligned}
$$

$$
\begin{aligned}
\delta_1(2) &= b_{1G} \max\{a_{11}\delta_1(1), a_{21}\delta_2(1)\} = 0.4 \max\{0.7 \cdot 0.24, 0.2 \cdot 0.08\} = 0.0672 \\
\Delta_1(2) &= \boxed{s_1} \\
\delta_2(2) &= b_{2G} \max\{a_{12}\delta_1(1), a_{22}\delta_2(1)\} = 0.2 \max\{0.3 \cdot 0.24, 0.8 \cdot 0.08\} = 0.0144 \\
\Delta_2(2) &= s_1
\end{aligned}
$$

$$
\begin{aligned}
\delta_1(3) &= b_{1C} \max\{a_{11}\delta_1(2), a_{21}\delta_2(2)\} = 0.1 \max\{0.7 \cdot 0.0672, 0.2 \cdot 0.0144\} = 0.004704 \\
\Delta_1(3) &= s_1 \\
\delta_2(3) &= b_{2C} \max\{a_{12}\delta_1(2), a_{22}\delta_2(2)\} = 0.3 \max\{0.3 \cdot 0.0672, 0.8 \cdot 0.0144\} = 0.006048 \\
\Delta_2(3) &= \boxed{s_1}
\end{aligned}
$$

$$
\begin{aligned}
\delta_1(4) &= b_{1T} \max\{a_{11}\delta_1(3), a_{21}\delta_2(3)\} = 0.1 \max\{0.7 \cdot 0.004704, 0.2 \cdot 0.006048\} = 0.00032928 \\
\Delta_1(4) &= s_1 \\
\delta_2(4) &= b_{2T} \max\{a_{12}\delta_1(3), a_{22}\delta_2(3)\} = 0.3 \max\{0.3 \cdot 0.004704, 0.8 \cdot 0.006048\} = \boxed{0.00145152} \\
\Delta_1(4) &= \boxed{s_2}
\end{aligned}
$$

$$\boxed{s^*_{1:4} = [s_1, s_1, s_2, s_2]}$$

4.3

$$
\begin{aligned}
P(O_5 = o_5 | O_{1:4}; \theta) &= P(O_5 = o_5, O_{1:4}) / P(O_{1:4}) \\
&= \sum_s P(O_5 = o_5, X_5 = s, O_{1:4}) / P(O_{1:4}) \\
&= \sum_s P(O_5 = o_5 | X_5 = s, \cancel{O_{1:4}}) P(X_5 = s, O_{1:4}) / P(O_{1:4}) \\
&= \sum_s b_{s,o_5} \sum_{s'} P(X_5 = s, X_4 = s', O_{1:4}) / P(O_{1:4}) \\
&= \sum_s b_{s,o_5} \sum_{s'} P(X_5 = s | X_4 = s', \cancel{O_{1:4}}) P(X_4 = s', O_{1:4}) / P(O_{1:4}) \\
&= \sum_s b_{s,o_5} \sum_{s'} a_{s',s} \alpha_{s'}(4) / P(O_{1:4})
\end{aligned}
$$

$$
\begin{aligned}
\implies o^* &= \operatorname{argmax}_{o_5} \sum_s b_{s,o_5} \sum_{s'} a_{s',s} \alpha_{s'}(4) / P(O_{1:4}) \\
&= \operatorname{argmax}_{o_5} \sum_s b_{s,o_5} \sum_{s'} a_{s',s} \alpha_{s'}(4)
\end{aligned}
$$

$$
\begin{aligned}
P(O_5 = A|O_{1:4} = [AGCT]) &= b_{1A}(a_{11}\alpha_1(4) + a_{21}\alpha_2(4)) + b_{2A}(a_{12}\alpha_1(4) + a_{22}\alpha_2(4)) \\
&= 0.4(0.7 \cdot 0.00066176 + 0.2 \cdot 0.00366912) + \\
&\quad\ 0.2(0.3 \cdot 0.00066176 + 0.8 \cdot 0.00366912) \\
&\approx \boxed{0.0011056} \\
P(O_5 = C|O_{1:4} = [AGCT]) &= b_{1C}(a_{11}\alpha_1(4) + a_{21}\alpha_2(4)) + b_{2C}(a_{12}\alpha_1(4) + a_{22}\alpha_2(4)) \\
&= 0.1(0.7 \cdot 0.00066176 + 0.2 \cdot 0.00366912) + \\
&\quad\ 0.3(0.3 \cdot 0.00066176 + 0.8 \cdot 0.00366912) \\
&\approx 0.0010598 \\
P(O_5 = G|O_{1:4} = [AGCT]) &= b_{1G}(a_{11}\alpha_1(4) + a_{21}\alpha_2(4)) + b_{2G}(a_{12}\alpha_1(4) + a_{22}\alpha_2(4)) \\
&= 0.4(0.7 \cdot 0.00066176 + 0.2 \cdot 0.00366912) + \\
&\quad\ 0.2(0.3 \cdot 0.00066176 + 0.8 \cdot 0.00366912) \\
&\approx \boxed{0.0011056} \\
P(O_5 = T|O_{1:4} = [AGCT]) &= b_{1T}(a_{11}\alpha_1(4) + a_{21}\alpha_2(4)) + b_{2T}(a_{12}\alpha_1(4) + a_{22}\alpha_2(4)) \\
&= 0.1(0.7 \cdot 0.00066176 + 0.2 \cdot 0.00366912) + \\
&\quad\ 0.3(0.3 \cdot 0.00066176 + 0.8 \cdot 0.00366912) \\
&\approx 0.0010598
\end{aligned}
$$

$$\boxed{o^* = A \text{ , or } o^* = G}$$