1. Consider the following two 2-state Hidden Markov Models where both states have two possible output symbols A and B.

Model 1:

Transition probabilities: \( a_{11} = 0.7, a_{12} = 0.3, a_{21} = 0.0, a_{22} = 1.0 \) (\( a_{ij} \) is the probability of going from state \( i \) to state \( j \))

Output probabilities: \( b_1(A) = 0.8, b_1(B) = 0.2, b_2(A) = 0.4, b_2(B) = 0.6 \)

Initial probabilities: \( \pi_1 = 0.5, \pi_2 = 0.5 \).

Model 2:

Transition probabilities: \( a_{11} = 0.6, a_{12} = 0.4, a_{21} = 0.0, a_{22} = 1.0 \)

Output probabilities: \( b_1(A) = 0.9, b_1(B) = 0.1, b_2(A) = 0.3, b_2(B) = 0.7 \)

Initial probabilities: \( \pi_1 = 0.4, \pi_2 = 0.6 \).

(a) Sketch the state diagram for two models.
(b) Which Model is more likely to produce the observation sequence \{A, B, B\}?  
(c) Given the observation sequence \{A, B, B\} find the Viterbi path for each model. Would your answer in (b) differ if you used the likelihood of the Viterbi paths to approximate the likelihood of the model?

2. Consider the following HMM:

![State Diagram]

\[
\begin{array}{c|ccc}
\text{Transition Probabilities} & a_{11} & a_{12} & a_{21} \\
\hline
\pi_1 & 1/2 & 1/2 & 0 \\
\pi_2 & 0 & 1/2 & 1/2 \\
\pi_3 & 0 & 0 & 1 \\
\end{array}
\]

\[
\begin{array}{c|ccc}
\text{Output Probabilities} & b_1(X) & b_1(Y) & b_1(Z) \\
\hline
b_2(X) & 1/2 & 1/2 & 0 \\
b_2(Y) & 1/2 & 0 & 1/2 \\
b_2(Z) & 1/2 & 1/2 & 0 \\
\end{array}
\]

\[
\begin{array}{c|ccc}
\text{Initial Probabilities} & \pi_1 & \pi_2 & \pi_3 \\
\hline
\pi_1 & 1/2 & 1/2 & 0 \\
\pi_2 & 0 & 1/2 & 1/2 \\
\pi_3 & 0 & 0 & 1 \\
\end{array}
\]

Where

\[
a_{ij} = P(q_{i-1} = S_j | q_i = S_j)
\]

\[
b_i(k) = P(O_i = k | q_i = S_j)
\]
Suppose we have observed this sequence: XZXYZZYZZ.

(In long-hand: \( O_1 = X, O_2 = Z, O_3 = X, O_4 = Y, O_5 = Z, O_6 = Z, O_7 = Y, O_8 = Z, O_9 = Z \).)

Fill in the following table with \( \alpha_i(t) \) values, remembering the definition:

\[
\alpha_i(t) = P(O_1 \land O_2 \land \ldots \land O_t \land q_t = s_t)
\]

So for example, \( \alpha_3(2) = P(O_1 = X \land O_2 = Z \land O_3 = X \land q_3 = S_2) \).

<table>
<thead>
<tr>
<th>t</th>
<th>( \alpha_1(1) )</th>
<th>( \alpha_1(2) )</th>
<th>( \alpha_1(3) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Warning: this is a question that will take a few minutes if you really understand HMMs, but could take hours if you don't!

3. (a) Working with linear SVMs:

Consider the following data set with one positive example \( x_1 = (0,0) \), \( y_1 = +1 \) and one negative example \( x_2 = (4,4) \), \( y_2 = -1 \).

Provide the SVM classifier below.

\[
h(x) = \begin{cases} 
  +1 & \text{if } \ldots x_1 + \ldots x_2 + \ldots \geq 0 \\
  -1 & \text{otherwise}
\end{cases}
\]
Also provide the weight value of the support vectors $\alpha_1$ and $\alpha_2$, as well as the offset threshold value $b$ of the SVM classifier.

(b) ON properties of linear SVMs I

Suppose we have an additional positive example $x_3 = (-1, -1)$, $y_3 = +1$.

(b-1) Which data points are support vectors? (Circle)

(b-2) Is the decision boundary for this data set different than for (a)?

(b-3) Relative to the support vectors for the data set in (a), the weights of the support vectors for the data set in this part are the same, smaller or larger?

(c) ON properties of linear SVMs II

Suppose we have an additional positive example $x_4 = (2, 2)$, $y_4 = +1$.

(c-1) Which data points are support vectors? (Circle)
(c-2) If both this data set and that in (b) are linearly separable, then the decision boundary for this data set different than for (b)?

(c-3) Relative to the support vectors for the data set in (b), the weights of the support vectors for the data set in this part are the same, smaller or larger?

(d) On power of SVM

Suppose we have an additional negative example \( x_5 = (-3, -3), y_5 = -1 \).

Circle the number(s) to the left side of the kernels described in the enumerated list below that can separate the data.

1. Linear: \( K(u, v) = u \cdot v \)

2. Polynomial of degree \( n \geq 2 \): \( K(u, v) = (1 + u \cdot v)^n \)

3. Radial Basis Function / Gaussian with sufficiently small scale parameter \( \sigma \): \( K(u, v) = e^{-\frac{|u-v|^2}{2\sigma^2}} \)

4. None

(e) Think of definition the new kernel

Suppose we have two additional examples: one positive example \( x_6 = (1, +1), y_6 = +1 \), and one negative example \( x_7 = (-4, -4), y_7 = -1 \).
Consider using the following kernel:

\[ K(u, v) = 2\|u\|\|v\| \]

Find the SVM classifier:

\[ h(x) = \begin{cases} 
+1 & \text{if } -2x_1^2 + \ldots + x_1^2 + \ldots x_2^2 + \ldots x_1^2 + \ldots \geq 0 \\
-1 & \text{otherwise} 
\end{cases} \]

Which data points are support vectors? (Circle)

4. There are the following nine diagrams, representing graphs of SVMs trained to separate pluses (+) from minuses (-). Indicate which diagram results from using which kernel function by writing the letter of the diagram next to the corresponding kernel. Note that the points in diagrams A, B, C and D are the same. And the points in E, F and G also are the same.

<table>
<thead>
<tr>
<th>( \kappa(u_1, u_2) = (u_1, u_2)^i )</th>
<th>( \kappa(u_1, u_2) = \exp\left(-\frac{|u_1 - u_2|^2}{0.5}\right) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \kappa(u_1, u_2) = (u_1, u_2)^2 )</td>
<td>( \kappa(u_1, u_2) = \exp\left(-\frac{|u_1 - u_2|^2}{0.22}\right) )</td>
</tr>
<tr>
<td>( \kappa(u_1, u_2) = (u_1, u_2)^2 )</td>
<td>( \kappa(u_1, u_2) = \exp\left(-\frac{|u_1 - u_2|^2}{0.08}\right) )</td>
</tr>
</tbody>
</table>