Date Page 1

1. Find a SVM classifier for the given data and classify the point (4,-4

X_1	X_2	class label
6	-2	+
12	2	+
6	2	+
12	-2	+
0	-2	-
-2	0	-
2	0	-
0	2	-

The support vectors can be

$S_1 = \begin{bmatrix} 6 \\ -2 \end{bmatrix}$ $S_2 = \begin{bmatrix} 6 \\ 2 \end{bmatrix}$ $S_3 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$

$y_1 = +1$ $y_2 = +1$ $y_3 = -1$

$\sum_{i=1}^{3} \alpha_i y_i = 0$

$\alpha_1(+1) + \alpha_2(+1) + \alpha_3(-1) = 0$

$\alpha_3 = \alpha_1 + \alpha_2$

$wx + h = -1 \Rightarrow ws_3 + b = -1$

$2w_1 + b = -1$ —— (1)

$wx + h = +1 \Rightarrow ws_1 + b = 1$

$6w_1 - 2w_2 + b = 1$ —— (2)

$ws_2 + b = 1$

$6w_1 + 2w_2 + b = 1$ —— (3)

(2) + (3) $\Rightarrow 12w_1 + 2b = 2$ —— (4)

(4) - 2(1) $\Rightarrow 8w_1 = 4$

$w_1 = \frac{1}{2} \Rightarrow b = -2, \; w_2 = 0$

$w = \begin{bmatrix} 1/2 \\ 0 \end{bmatrix}$ $b = -2$

$$w = \begin{bmatrix} 1/2 \\ 0 \end{bmatrix} \qquad b = -2$$

give $x = [4, -4]$

$$wx + b = 4(1/2) - 4(0) - 2$$

$$= 2 - 2$$

$$= 0$$

$$wx + b = 0$$

2 consider the following 2D data points: Find out the points that belongs to each cluster by assuming K=3 and A4, A8 and A12 are the cluster centers of C1, C2 and C3 respectively. Perform 2 iterations.

Point	Co-ordinates
A1	(2,10)
A2	(2,6)
A3	(11,11)
A4	(6,9)
A5	(6,4)
A6	(1,2)
A7	(5,10)
A8	(4,9)
A9	(10,12)
A10	(7,5)
A11	(9,11)
A12	(4,6)
A13	(3,10)
A14	(3,8)
A15	(6,11)

Iteration 1

Data points	Distance to C1	Distance to C2	Distance to C3
A1	6.08	3.16	4.47
A2	4.12	2.24	2.83
A3	8.06	2.24	5.10
A4	0	3.61	5.10
A5	5.66	2.24	3.61
A6	7.07	5.83	2.24
A7	3.16	0.41	5.83

A8	2.24	0	4.12
A9	5.39	6.71	3.16
A10	5.83	4.47	2.24
A11	4.12	0	3.16
A12	2.24	3.61	0
A13	4.12	0.41	4.47
A14	4.47	1.41	5.10
A15	5.37	0	3.16

Data point	Assigned cluster
A1	C3
A2	C3
A3	C2
A4	C1
A5	C2
A6	C3
A7	C1
A8	C2
A9	C2
A10	C2
A11	C2
A12	C3
A13	C3
A14	C3
A15	C2

cluster	New center
C1	(5.5, 9.5)
C2	(6.6, 6.2)
C3	(4, 6)

Iteration 2

Data Point	Distance to C1	Distance to C2	Distance to C3
A1	6.92	3.20	3.16

A2	4.58	2.32	2
A3	9.11	2.32	4.47
A4	2.83	3.88	4.47
A5	4.74	2.32	3.88
A6	7.81	6.10	1.41
A7	2	0.89	5.83
A8	0.89	0	3.16
A9	4.58	6.79	2.32
A10	5.39	4.47	1.41
A11	2.83	0	2.32
A12	0.89	3.16	0
A13	3.16	0.89	2.32.
A14	3.6		

~~A15~~

Data point	Assigned cluster
A1	C3
A2	C3
A3	C2
A4	C1
A5	C2
A6	C3
A7	C1
A8	C2
A9	C3
A10	C3
A11	C2
A12	C3
A13	C2
A14	C1

cluster	center
C1	(11, 11)
C2	(6, 6)
C3	(3, 3)

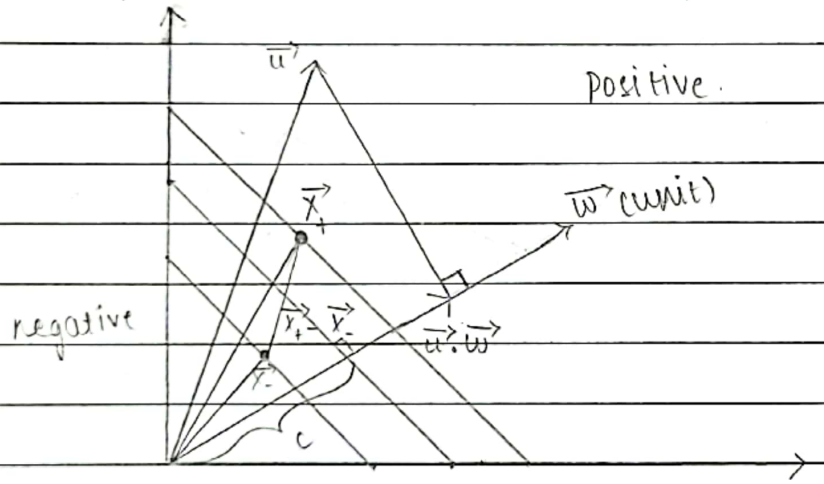
3. The SVM hard margin classifier aims to find the optimal hyperplane that maximizes the margin between the two classes while ensuring that all training points are correctly classified. Derive the following:

a. The optimization problem that needs to be solved for finding the optimal hyperplane.

b. The lagrangian for the optimization problem.

c. The dual form of the optimization problem.

d. The equations for w and b in terms of the lagrange multipliers.



(a) In other words,

$\vec{w}.\vec{u} \geq c$ — positive

$\vec{w}.\vec{u} \leq c$ — negative

$\vec{w}.\vec{u} - c \geq 0$ — (1)

$\vec{w}.\vec{u} - c \leq 0$ — (2)

Let $\vec{u} = \vec{x_+}$ for positive sample and $\vec{x_-}$ for negative sample

At the boundary, let the 2 constants be c_1 and c_2

$\vec{w}.\vec{x_+} - c_1 = 0$ — (3) $\quad \vec{w}\,\vec{x_-} - c_2 = 0$ — (4)

Let $A = -\frac{(c_2 + c_1)}{2}$ and $B = -\frac{(c_2 - c_1)}{2}$

Then $-C_1 = A - B$ and $C_2 = A + B$

Replacing in (3) and (4)

$\vec{w}\,\vec{x}_+ + A - B = 0$ — (5) $\qquad \vec{w}\,\vec{x}_- + A + B = 0$ — (6)

Dividing (5) and (6) by B

$\frac{\vec{w}}{B}\,\vec{x}_+ + \frac{A}{B} - 1 = 0$ — (7) $\qquad \frac{\vec{w}}{B}\,\vec{x}_- + A + 1 = 0$ — (8)

Redefining $\frac{\vec{w}}{B} = \vec{w}$ and $\frac{A}{B} = b$ in (7) and (8)

$\vec{w}\,\vec{x}_+ + b = 1$ — (9)

$\vec{w}\,\vec{x}_- + b = -1$ — (10)

converting to inequality

$\vec{w}\,\vec{x}_+ + b \geq 1$ — (11)

$\vec{w}\,\vec{x}_- + b \leq -1$ — (12)

The predicted values of y

$y_i = 1$: +ve

$y_i = -1$: -ve

multiplying (11) and (12) with y_i

$y_i(\vec{w}.\vec{x}_+ + b) - 1 \geq 0$ — (13)

$y_i(\vec{w}.\vec{x}_- + b) - 1 \leq 0$ — (14)

generalizing for all $\vec{x}$ s

$y_i(\vec{w}.\vec{x} + b) - 1 \geq 0$ — (15)

(b) Multiple constraints

$L(x, \lambda) = f(x) + \lambda_1 g_1(x) + \lambda_2 g_2(x) + \ldots\ldots + \lambda_k g_k(x)$

Objective $f(w) = \frac{1}{2}\|w\|^2$ to minimize

Subject to m constraints $g(w, b) = y_i(w^T x_i + b) - 1$

$L = f(w) - \sum_{i=1}^{m} \alpha_i g_i(w, b)$

$$L = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{M} \alpha_i [y_i (w^T x_i + b) - 1] \quad \text{——— (16)}$$

(d) Derivative wrt w

$$\frac{\partial L}{\partial w} = \vec{w} - \Sigma \alpha_i y_i \vec{x_i} = 0$$

$$\vec{w} = \sum_{i=1}^{M} \alpha_i y_i \vec{x_i} \quad \text{——— (17)}$$

Derivative wrt b

$$\frac{\partial L}{\partial w} = -\sum_{i=1}^{M} \alpha_i y_i = 0$$

$$\sum_{i=1}^{M} \alpha_i y_i = 0 \quad \text{——— (18)}$$

(c) Replacing (17) and (18) in (16)

$$L = \frac{1}{2}\left(\sum_{i=1}^{M}\sum_{j=1}^{M} \alpha_i \alpha_j y_i y_j x_i^T x_j\right) - \left(\sum_{i=1}^{M}\sum_{j=1}^{M} \alpha_i \alpha_j y_i y_j x_i^T - x_j\right) - b\sum_{i=1}^{M} \alpha_i y_i$$

$$+ \sum_{i=1}^{M} \alpha i$$

$$L(w, b, \alpha) = \sum_{i=1}^{M} \alpha_i - \frac{1}{2}\left(\sum_{i=1}^{M}\sum_{j=1}^{M} \alpha_i \alpha_j y_i y_j x_i^T x_j\right)$$

$$\max_{\alpha} W(\alpha) = \sum_{i=1}^{M} \alpha_i - \frac{1}{2}\left(\sum_{i=1}^{M}\sum_{j=1}^{M} \alpha_i \alpha_j y_i y_j x_i^T x_j\right)$$

4. Suppose you have a 40% chance of it raining (event A) and there is a 60% chance that it will be cloudy (event B) on any given day. You also know that when it rains, there's a 45% chance of it being cloudy.

Bayes theorem $P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$

Here Event A is raining.
Event B is being cloudy.

P(A) - probability of rain 40% or 0.4
P(B) - probability of being cloudy 60% or 0.6
P(B|A) - probability of being cloudy given that it is raining 45% or 0.45

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

$$= \frac{0.45 \times 0.4}{0.6}$$

$$P(A|B) = 0.3$$

∴ probability that it is raining given that it is cloudy is 30%.

5. Suppose you have two classes, class A and class B, and two features (X and Y). For class A, the conditional probabilities of the features are as follows:

$P(X=1|A) = 0.2$ $\quad$ $P(Y=1|A) = 0.3$

For class B, the conditional probabilites of the features are

$P(X=1|B) = 0.8$ $\quad$ $P(Y=1|B) = 0.7$

The prior probabilites of the classes are:

$P(A) = 0.3$ $\quad$ $P(B) = 0.7$

You're given a new instances with features $X=1$ and $Y=1$. use the Baye's optimal classifier to determine which class is most probable for this instance.

for $X=1$ $\quad$ $Y=1$

$$P(A|X=1, Y=1) = \frac{P(X=1, Y=1|A) \cdot P(A)}{P(X=1, Y=1)}$$

$$P(B|X=1, Y=1) = \frac{P(X=1, Y=1|B) \cdot P(B)}{P(X=1, Y=1)}$$

- For class A

$P(X=1, Y=1|A) = P(X=1|A) \cdot P(Y=1|A)$

$= 0.2 \times 0.3$

$P(X=1, Y=1|A) = 0.06$

$P(A|X=1, Y=1) \propto 0.06 \times 0.3$

$P(A|X=1, Y=1) = 0.018$

- For class B

$P(B|X=1, Y=1|B) = P(X=1|B) \cdot P(Y=1|B)$

$= 0.8 \times 0.7$

$P(X=1, Y=1|B) = 0.56$

$P(B|X=1, Y=1) \propto 0.56 \times 0.7$

$P(B|X=1, Y=1) = 0.392$

Normalizing the probabilities:

$$P(A|X=1, Y=1) = \frac{0.018}{0.018 + 0.392} = \frac{0.018}{0.41} = 0.0439$$

$$P(B|X=1, Y=1) = \frac{0.392}{0.018 + 0.392} = \frac{0.392}{0.41} = 0.956$$

According to the Bayes optimal classifier the new instance with features $X=1$ and $Y=1$ is most probable to belong to class B

6. consider the following data:

Text	class
It was a bad movie	-
the movie had no plot	-
It was a great movie	+
the movie had a good plot	+
the movie was a boring movie	-

classify the text "the movie had a boring plot"

Note: consider stop words for this.

The text "the movie had a boring plot" can be classified as negative.

(i) Remove stop words: stop words are words that are common in a language and do not add much meaning to a text. In this case the stop words are 'the', 'a', and had. After removing the stop words, the text is "movie boring plot".

(ii) The word boring has a negative sentiment, so sentiment of the text is negative.

∴ the text "The movie had a boring plot" can be classified as negative.

7. given the following states and their transition and emission probabilities, calculate the probability of the output sequence
$O1 \rightarrow O3 \rightarrow O2$ using the forward algorithm by using the given HMM parameters

$\pi A = 0.3$ $\pi B = 0.2$ $\pi C = 0.5$

Transition and emission table:

	A	B	C	O1	O2	O3
A	0.1	0.5	0.4	0.2	0.2	0.6
B	0.7	0.2	0.1	0.4	0.4	0.2
C	0.1	0.1	0.8	0.5	0.4	0.1

where O1, O2, O3 are the emission states.

(i) Initialize the forward probabilites:

$\alpha 1(A) = \pi A \times P(O1|A)$

$= 0.3 \times 0.2$

$\alpha 1(A) = 0.06$

$\alpha 1(B) = \pi B \times P(O1|B)$

$= 0.2 \times 0.4$

$\alpha 1(B) = 0.08$

$\alpha 1(C) = \pi C \times P(O1|C)$

$= 0.5 \times 0.5$

$\alpha 1(C) = 0.25$

(ii) calculate the forward probabilites for time step $t = 2$

$\alpha 2(A) = \alpha 1(A) \times P(B|A) \times P(O2|B)$

$= 0.06 \times 0.5 \times 0.4$

$\alpha 2(A) = 0.012$

$\alpha 2(B) = \alpha 1(A) \times P(C|B) \times P(O2|C)$

$= 0.08 \times 0.1 \times 0.4$

$\alpha 2(B) = 0.0032$

$\alpha_2(C) = \alpha_1(C) \times P(A|C) \times P(O_2|A)$

$= 0.25 \times 0.8 \times 0.2$

$\alpha_2(C) = 0.04$

(3) Calculate the forward probabilites for time step $t=3$:

$\alpha_3(A) = \alpha_2(A) \times P(C|A) \times P(O_3|C)$

$= 0.012 \times 0.4 \times 0.6$

$\alpha_3(A) = 0.00288$

$\alpha_3(B) = \alpha_2(B) \times P(A|B) \times P(O_3|A)$

$= 0.0032 \times 0.7 \times 0.2$

$\alpha_3(B) = 0.000448$

$\alpha_3(C) = \alpha_2(C) \times P(B|C) \times P(O_3|B)$

$= 0.04 \times 0.1 \times 0.2$

$\alpha_3(C) = 0.0008$

(4) Calculate the probability of the output sequence

$P(O_1 \rightarrow O_3 \rightarrow O_2) = \alpha_3(A) + \alpha_3(B) + \alpha_C(C)$

$= 0.00288 + 0.000448 + 0.0008$

$P(O_1 \rightarrow O_3 \rightarrow O_2) = 0.004128$

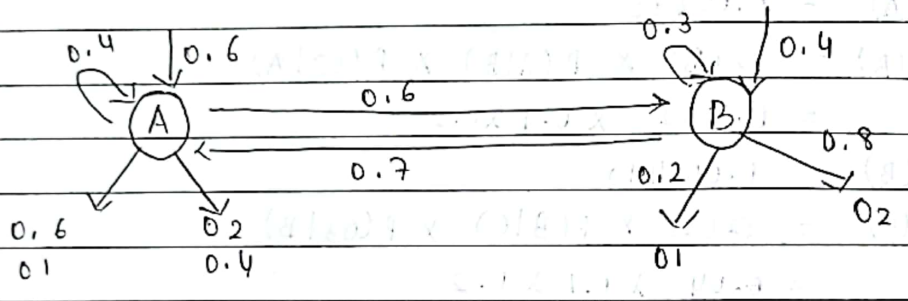
∴ The probability of the output sequence $O_1 \rightarrow O_3 \rightarrow O_2$ is 0.004128.

8. Given the following HMM, use Viterbi's algorithm to compute the most probable final hidden state given the following output sequence is generated:

Output sequence:

$O_2 \rightarrow O_1 \rightarrow O_1$

HMM:



(1) Initialize the viterbi probabilities

$\delta_1(A) = \pi A \times P(O_2|A)$

$= 0.3 \times 0.2$

$\delta_1(A) = 0.06$

$\delta_1(B) = \pi B \times P(O_2|B)$

$= 0.2 \times 0.4$

$\delta_1(B) = 0.08$

$\delta_1(C) = \pi C \times P(O_2|C)$

$= 0.5 \times 0.1$

$\delta_1(C) = 0.05$

(2) Calculate the viterbi probabilites and back pointers for time step $t = 2$

$\delta_2(A) = \max\{\delta_1(A) \times P(A|A) \times P(O_1|A), \delta_1(B) \times P(B|A) \times P(O_1|B), \delta_1(C) \times P(C|A) \times P(O_1|C)\}$

$= \max\{0.6 \times 0.1 \times 0.2, 0.08 \times 0.7 \times 0.4, 0.05 \times 0.8 \times 2.4\}$

$\delta_2(A) = 0.0224$

$\delta_2(B) = \max\{\delta_1(A) \times P(A|B) \times P(O_1|A), \delta_1(B) \times P(B|B) \times P(O_1|B),$

$\delta_1(C) \times P(C|B) \times P(O1|C)\}$ a

$= \max\{0.06 \times 0.5 \times 0.2, 0.8 \times 0.2 \times 0.4, 0.05 \times 0.8 \times 0.5\}$

$= 0.0192$

$\delta_2(C) = \max\{\delta_1(A) \times P(A|C) \times P(O1|A), \delta_1(B) \times P(B|C) \times P(O1|B),$
$\delta_1(C) \times P(C|C) \times P(O1|C)\}$

$= \max\{0.06 \times 0.8 \times 0.2, 0.08 \times 0.1 \times 0.4, 0.05 \times 0.8 \times 0.5\}$

$= 0.02$

(3) calculate the viterbi probabilities and back pointers for time step $t = 3$

$\delta_3(A) = \max\{\delta_2(A) \times P(A|A) \times P(O1|A), \delta_2(B) \times P(B|A) \times P(O1|B),$
$\delta_2(C) \times P(C|A) \times P(O1|C)\}$

$= \max\{0.0224 \times 0.1 \times 0.2, 0.0192 \times 0.7 \times 0.4, 0.02 \times 0.8 \times 0.2\}$

$\delta_3(A) = 0.00576$

$\delta_3(B) = \max\{\delta_2(A) \times P(A|B) \times P(O1|A), \delta_2(B) \times P(B|B) \times P(O1|B),$
$\delta_2(C) \times P(C|B) \times P(O1|C)\}$

$\delta_3(B) = \max\{0.0224 \times 0.5 \times 0.2, 0.0192 \times 0.2 \times 0.4, 0.02 \times 0.1 \times 0.2\}$

$\delta_3(B) = 0.00448$

$\delta_3(C) = \max\{\delta_2(A) \times P(A|C) \times P(O1|A), \delta_2(B) \times P(B|C) \times P(O1|B),$
$\delta_2(C) \times P(C|C) \times P(O1|C)\}$

$= \max\{0.0224 \times 0.8 \times 0.2, 0.0192 \times 0.1 \times 0.4, 0.02 \times 0.8 \times 0.5\}$

$\delta_3(C) = 0.00768.$

(4) The most probable final hidden state is the state with the higher viterbi probability at time step $t = 3$. In this case, the most probable final hidden state is C, with a viterbi probability of 0.00768.

9 Suppose we have a chicken from which we collect eggs at noon every day. Now whether or not the chicken has laid eggs for collection depends on some unknown factors that are hidden we can however assume that the chicken is always is one of 2 states that influence whether the chicken lays eggs, and that this state only depends on the state on the previous day. Now we don't know the state at the initial starting point, we don't know the transition probabilities between the 2 states and we don't know the probability that the chicken lays an egg given a particular state. To start we first guess the transition and emission matrices

Transition

	State 1	State 2
State 1	0.5	0.5
State 2	0.3	0.7

Emission

	No eggs	Eggs
State 1	0.3	0.7
State 2	0.8	0.2

Initial

State 1	0.2
State 2	0.8

We then take a set of observations (E = eggs ; N = no eggs):

N, N, N, N, N, E, E, N, N, N

This gives us a set of observed transition between days:

NN, NN, NN, NN, NE, EE, EN, NN, NN

using the above information, compute the values in the following tables:

Observed sequence	Highest probability for S2 > S1	Highest probability of observing that sequence	
		Probs Here	Sequence here
NN	~~S1 -> S2~~ –	0.3804	S2, S2
NNN	~~S1 -> S2~~ –	0.2408	S2, S2, S2
NNNN	~~S1 -> S2~~ –	0.1526	S2, S2, S2, S2
NNNNN	~~S1 -> S2~~ –	0.0968	S2, S2, S2, S2, S2

NNNNNE	5th N to 1st E	Highest S2 → S1	0.0153	S2,S2,S2,S2,S2,S1
NNNNNEE		Highest S2 → S1	0.0048	S2,S2,S2,S2,S2,S1,S1
NNNNNEEN		Highest S2 → S1	0.0077	S2,S2,S2,S2,S2,S1,S1,S2
NNNNNEENN		Highest S2 → S1	0.0052	S2,S2,S2,S2,S2,S1,S1,S2,S2
NNNNNEENNN		S1 → S2	0.00398	S2,S2,S2,S2,S2,S1,S1,S2,S2,S2
N		S1 → S2		
Total		S1 → S1		0.00398

New emission matrix estimate:

Observed sequence	Highest probability of observing the sequence if E assumed to come from S2.		Highest probability of observing that sequence	
	probs here	sequence here	probs here	Sequence here
NN	0 (E never comes from S2)		0.3804	S2,S2
NNN	0 (E never comes from S2)		0.2408	S2,S2,S2
NNNN	0 (E never comes from S2)		0.1526	S2,S2,S2,S2
NNNNN	0 (E never comes from S2)		0.0968	S2,S2,S2,S2,S2
NNNNEE	0 (E never comes from S2)		0.0153	S2,S2,S2,S2,S2,S1,S1
NNNNEEN	0 (E never comes from S2)		0.0048	S2,S2,S2,S2,S2,S1,S1
NNNNEENN	0 (E never comes from S2)		0.0077	S2,S2,S2,S2,S2,S1,S1,S2
NNNNEENNN	0 (E never comes from S2)		0.0052	S2,S2,S2,S2,S2,S1,S1,S2,S2
NNNNEENNN	0 (E never comes from S2)		0.00398	S2,S2,S2,S2,S2,S1,S1,S2,S2,S2
Total	0			0.00398

10. State the following algorithms in terms of initialization, induction / recursion and termination coiteria for HMM in terms of alpha, beta, gamma and di-gamma probabilities Please define symbols for each and draw a diagram for the induction process of the following algorithms:

(i) Forward Algorithm

(ii) Backward Algorithm

(iii) Viterbi Algorithm

(i) Forward Algorithm

compute likelihood till the $t-1$ timestamp

update likelihood at time t based on observation O_t

Let $\alpha_t(j)$ = joint probability of observing observations $O_1, O_2 \ldots O_t$ and reaching state q_j at time t.

$$\alpha_t(j) = P(O_1, O_2, \ldots O_t, q_t = j | \lambda)$$

$$= \sum_{i=1}^{N} P(O_1, O_2, \ldots O_t, q_{t-1} = i, q_t = j | \lambda)$$

Simplifying $O_1, O_2, \ldots O_{t-1} = O_{1,t-1}$

$$= \sum_{i=1}^{N} P(O_{1,t-1}, q_{t-1} = i) P(q_t = j | q_{t-1} = i, O_{1,t-1}) P(O_t | q_t = j, q_{t-1} = i, O_{1,t-1})$$

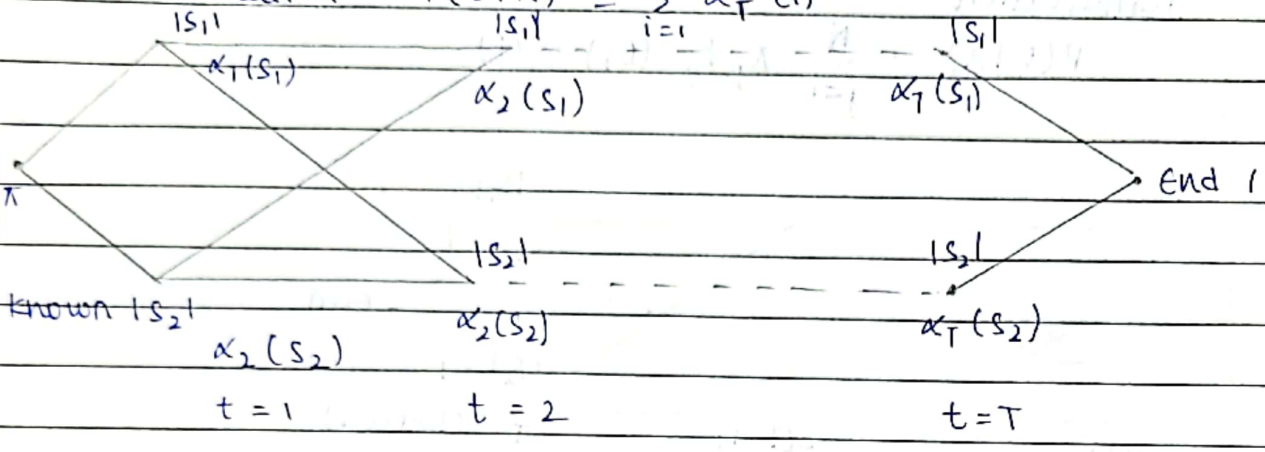
$$= \sum_{i=1}^{N} P(O_{1,t-1}, q_{t-1} = i) \underbrace{P(q_t = j | q_{t-1} = i)}_{\text{independent of } O_{1,t-1}} \underbrace{P(O_t | q_t = j, q_{t-1} = i)}_{\text{independent of } O_{1,t-1}}$$

$$\alpha_t(j) = \sum_{i=1}^{N} \alpha_{t-1}(i)\, a_{ij}\, b_j(O_t)$$

Recursion $\alpha_t(j) = \sum_{i=1}^{N} \alpha_{t-1}(i)\, a_{ij}\, b_j(O_t)$

Initialization $\alpha_1(i) = \pi_i\, b_i(O_1)$

Termination $P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$



(b) Backward Algorithm

likelihood of observing from t+1 to T

Let $\beta_t(j)$ be probability that future observations $O_{t+1}, O_{t+2}, \ldots O_T$ have been observed given that hidden state at t is j

$$\beta_t(j) = P(O_{t+1}, O_{t+2}, \ldots O_T | q_t = j, \lambda)$$

$$\beta_t(j) = \sum_{i=1}^{N} P(O_{t+1}, O_{t+2}, \ldots O_T, q_{t+1} = i | q_t = j, \lambda)$$

$$\beta_t(j) = \sum_{i=1}^{N} P(O_{t+1}, O_{t+2}, T, q_{t+1} = i | q_t = j, \lambda)$$

$$\beta_t(j) = \sum_{i=1}^{N} P(O_{t+1}, O_{t+2}, T | q_{t+1} = i, q_t = j, \lambda) P(q_{t+1} = i | q_t = j, \lambda)$$

$$\beta_t(j) = \sum_{i=1}^{N} P(O_{t+1}, O_{t+2}, T | q_{t-1} = i, q_t = j, \lambda) a_{ji}$$

$$\beta_t(j) = \sum_{i=1}^{N} P(O_{t+2,T} | O_{t+1}, q_{t+1} = i, q_t = j, \lambda) P(O_{t+1} | q_{t+1} = i, q_t = j) a_{ji}$$

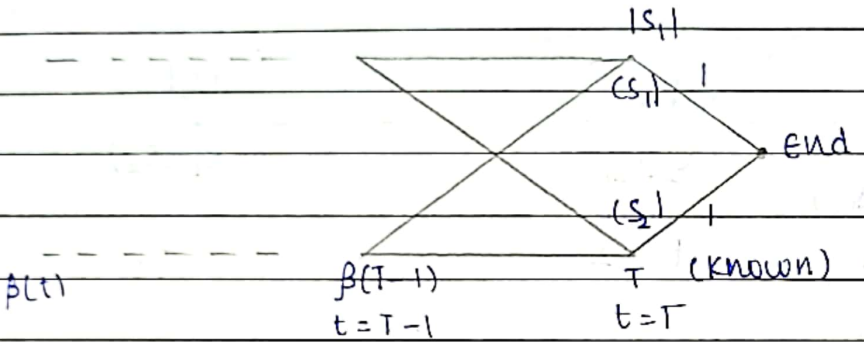
$$\beta_t(j) = \sum_{i=1}^{N} \beta_{t+1}(i) b_i(O_{t+1}) a_{ji}$$

Initialization $\beta_T(i) = 1$

Recursion $\beta_T(i) = \sum_{i=1}^{N} a_{ij} b_j(O_{t+1}) (\beta_{t+1}(j))$

Termination

$$P(O|\lambda) = \sum_{j=1}^{N} \pi_j b_j(O_1) \beta_1(j)$$



(c) Viterbi Algorithm

$$V_t(j) = \max_{q_1, q_2 \cdots q_{t-1}} P(q_1, q_2, \ldots q_{t-1}, O_1, O_2, \ldots O_{t-1}, q_t = j, O_t | \lambda)$$

Initialization

$$V_1(j) = \pi_j b_j(O_1)$$

$$bt_1(j) = 0$$

Recursion

$$V_t(j) = \max_{i=1}^{N} V_{t-1}(i)\, a_{ij}\, b_j(O_t)$$

$$bt_t(j) = \arg\max_{i=1}^{N} V_{t-1}(i)\, a_{ij}\, b_j(O_t)$$

Termination

Best Score $P^* = \max_{i=1}^{N} V_T(i)$

Best path start : $q_{T^*} = \arg\max_{i=1}^{N} V_T(i)$