CS461 – RECITATION 10

MACHINE LEARNING PRINCIPLES

Daize Dong

2025-11-17

TODAY’S CONTENT

- Quiz 04
- Homework 04

QUIZ 04

QUIZ 04

A	B	C	D
2	3	16	-

1. **When comparing a HMM to other ML techniques from this course, which is true?**

(a) HMMs can be trained with the EM algorithm, similar to GMMs and K-Means.

(b) HMMs capture temporal relationships between observations of arbitrary length, which is difficult to do with the other techniques so far covered in this class.

(c) Both of the above are true.

(d) All of the above are false.

EM in HMM: forward-backward.

QUIZ 04

A	B	C	D
3	1	15	-

2. When is a HMM least effective at predicting hidden states given observed states?

(a) When there are no state transitions (meaning each state transitions to itself with probability 1).

(b) When each state's transition probability to remain in the same state is very high, but less than 1.

(c) When state transitions are all equally likely. Lost all information. (uniform distribution)

(d) A HMM is always effective at predicting hidden states.

QUIZ 04

A	B	C	D
18	2	1	-

3. What is true about the stationary distribution, π, of a hidden Markov model?

(a) π= π A.

(b) Unlike the initial state distribution, the stationary distribution does not need to sum to 1.

(c) π= π ABO.

(d) None of the above.

Basic definition.

QUIZ 04

A	B	C	D
-	-	5	14

4. Which statement best describes the kernel trick? implicitly represent polynomial features

(a) Fitting polynomial parameters will better capture the variance of a dataset.

(b) A perceptron's Gram matrix is multiplied by a kernel function, improving the perceptron. replace inner product

(c) The kernel trick requires O(n^3) additional computation. can be any

(d) Linear parameters can be mapped onto non-linear spaces, increasing the discriminative power of linear models.

QUIZ 04

A	B	C	D
17	1	1	-

5. Which of these statements about a hidden Markov model with 3 hidden states and a vocabulary of 5 observable tokens is true?

(a) The parameters of the HMM can be trained with the EM algorithm.

(b) The transition matrix, A, is size 3x5. 5x5

(c) The emission matrix, B, is size 5x5. 3x5

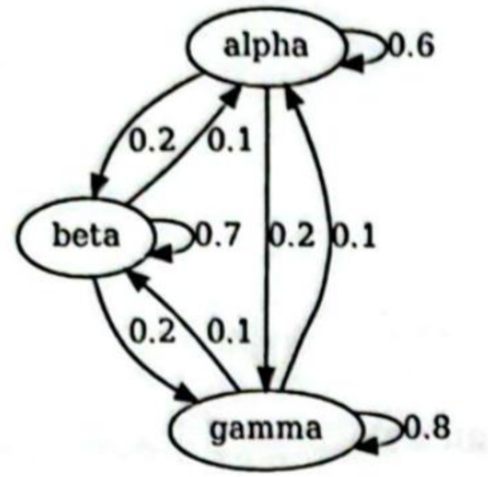
(d) b and C are both true.

Transition matrix: probabilities transiting to the next state.

Emission matrix: probabilities of all states during different time steps.

QUIZ 04

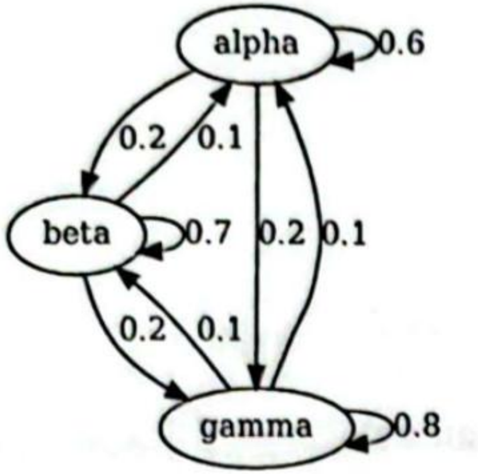
6. You are playing a popular online game and would like to maximize the time you spend having fun, which is a great use case for HMMs. There are three servers for the online game, named α, β, and γ, each with their own packet loss characteristics. You do not know which server you connect to, but you suspect that one of them is of better quality than the other two. You begin logging data, binning observations of packet loss over 1 minute intervals. You decide to assign packet loss statistics into three groups: **low, medium,** and **high**. After some sleuthing, you learn that users are transitioned from server to server as shown in the figure on the other side of the page.



QUIZ 04

Part 1: Populate a state transition matrix, A, using the values in the figure below.

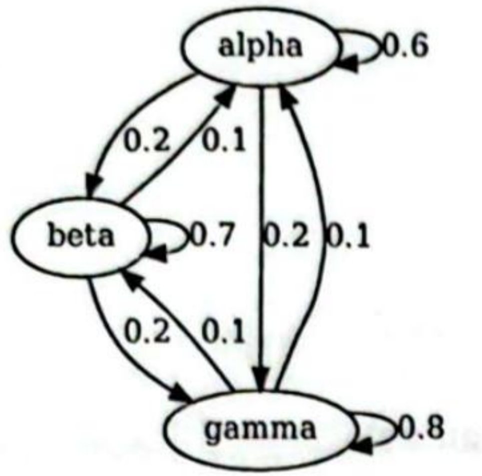
$$
A = \begin{array}{c c}
 & \begin{array}{ccc} \alpha & \beta & \gamma \end{array} \\
\begin{array}{c} \alpha \\ \beta \\ \gamma \end{array} & \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.1 & 0.7 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}
\end{array}
$$



QUIZ 04

Part 2: You learn that α is the best server, with low packet loss 2/3 of the time and medium the other 1/3. β's packet loss is medium half of the time and is evenly split between low and high the other half, while γ has medium packet loss 1/2 of the time and high packet loss the other 1/2. **Populate the emission matrix, B.**

$$B = \begin{array}{c} \\ \alpha \\ \beta \\ \gamma \end{array} \begin{array}{c} \begin{array}{ccc} \text{low} & \text{mid} & \text{high} \end{array} \\ \begin{bmatrix} 2/3 & 1/3 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 1/2 & 1/2 \end{bmatrix} \end{array}$$



QUIZ 04

Part 3: To prevent players from logging of instantly, the game always places users with server α for the first minute. In the second minute, you observe a medium packet loss. **Use $\pi_0 ABO$ to calculate the probabilities of being on each server α, β, and γ.**

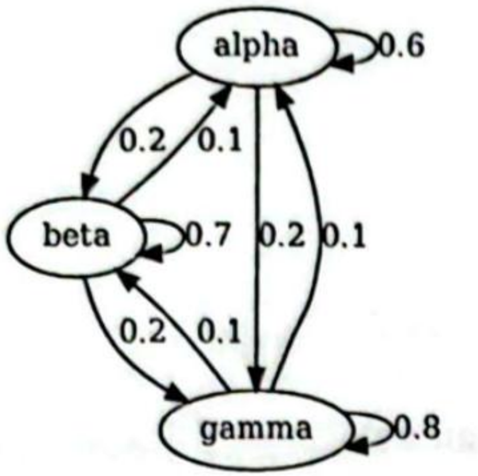
$$\pi_0 = \begin{bmatrix}1\\0\\0\end{bmatrix}^T, O = \begin{bmatrix}0\\1\\0\end{bmatrix}$$

$$\pi_{1|O} = \pi_0 A \odot BO$$

$$= \begin{bmatrix}1\\0\\0\end{bmatrix}^T \begin{bmatrix}0.6 & 0.2 & 0.2\\0.1 & 0.7 & 0.1\\0.1 & 0.1 & 0.8\end{bmatrix} \odot \begin{bmatrix}2/3 & 1/3 & 0\\1/4 & 1/2 & 1/4\\0 & 1/2 & 1/2\end{bmatrix}\begin{bmatrix}0\\1\\0\end{bmatrix}^T$$

$$= \begin{bmatrix}0.6\\0.2\\0.2\end{bmatrix}^T \odot \begin{bmatrix}1/3\\1/2\\1/2\end{bmatrix}^T$$

$$= \begin{bmatrix}0.2\\0.1\\0.1\end{bmatrix}^T \xRightarrow{\text{normalize}} \begin{bmatrix}0.5\\0.25\\0.25\end{bmatrix}^T$$



QUIZ 04

Part 4: On the first day of winter break, you decide to do nothing but play this game. You are ecstatic whenever you are using server α and frustrated when using the other servers. **What fraction of the time will you spend being ecstatic?**

This is to say what is the probability for α in stationary distribution.

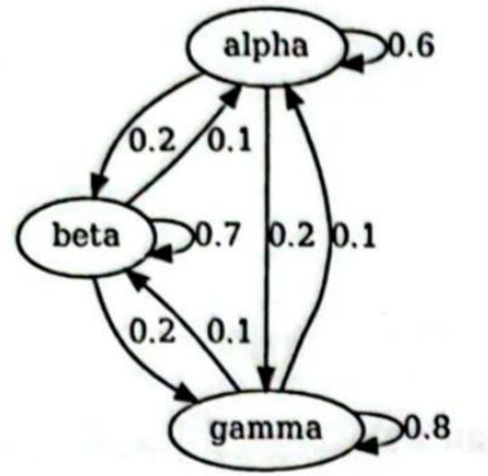
$\pi = \pi A$ **and** $\sum_{s \in S} \pi(s) = 1$

$$\pi = \pi A \rightarrow \begin{bmatrix} \pi(\alpha) \\ \pi(\beta) \\ \pi(\gamma) \end{bmatrix}^T = \begin{bmatrix} \pi(\alpha) \\ \pi(\beta) \\ \pi(\gamma) \end{bmatrix}^T \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.1 & 0.7 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$$

$$\rightarrow \begin{cases} \pi(\alpha) = 0.6\pi(\alpha) + 0.1\pi(\beta) + 0.1\pi(\gamma) \\ \pi(\beta) = 0.2\pi(\alpha) + 0.7\pi(\beta) + 0.1\pi(\gamma) \\ \pi(\gamma) = 0.2\pi(\alpha) + 0.1\pi(\beta) + 0.8\pi(\gamma) \end{cases} \textbf{and } \pi(\alpha) + \pi(\beta) + \pi(\gamma) = 1$$

$$\rightarrow \begin{cases} \pi(\alpha) = 0.2 \\ \pi(\beta) = 0.5 \\ \pi(\gamma) = 0.3 \end{cases}$$

So it's 20% time.



HOMEWORK 04

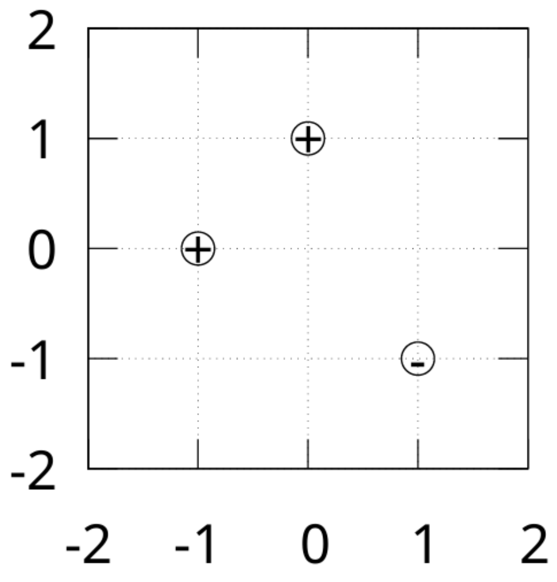
HOMEWORK 04

First, remember that we choose alpha to maximize an equation:

$$\underset{\alpha}{\operatorname{argmax}} \quad -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^T x_j + \sum_{i=1}^{n}\alpha_i$$

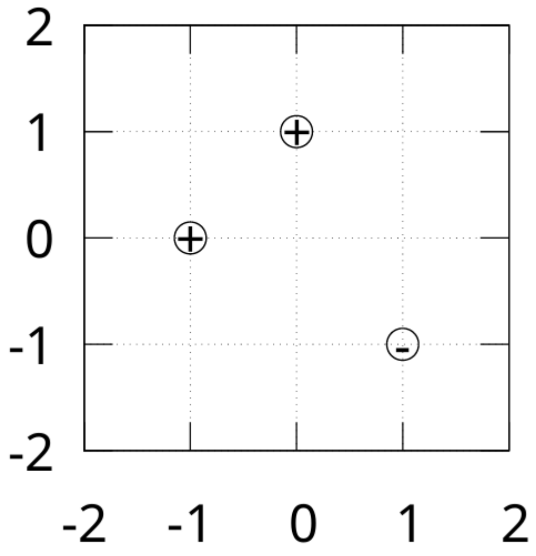
In addition, we have the following two constraints:
$\sum_{i=1}^{N}\alpha_i y_i = 0$
$0 \leq \alpha_i \forall i \in N$

<u>Part 1</u>: Given $X = [(0,1), (-1,0), (1,-1)]$ and $Y = (1,1,-1)$, solve for $\alpha_1, \alpha_2, \alpha_3$. Note that the two points from class 1 are symmetrical relative to the third point, so $\alpha_1 = \alpha_2$.



HOMEWORK 04

$$\underset{\alpha}{\text{argmax}} \quad -\tfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^T x_j + \sum_{i=1}^{n}\alpha_i$$

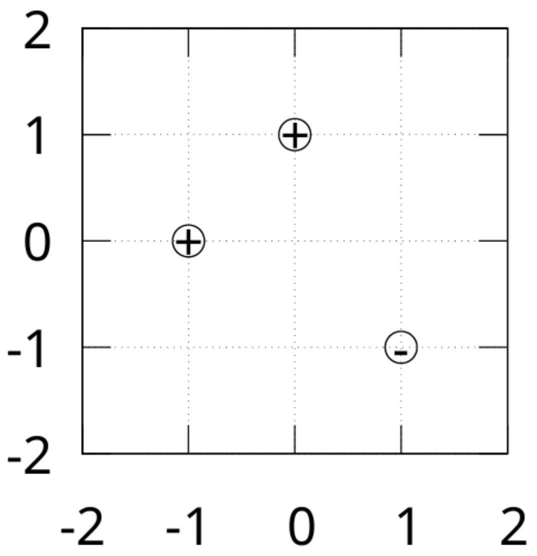


$$
\begin{aligned}
&= -\frac{1}{2}\alpha_1^2(1) - \frac{1}{2}\alpha_2^2 - \frac{1}{2}\alpha_3^2(2) - \alpha_1\alpha_2(0) + \alpha_1\alpha_3(-1) + \alpha_2\alpha_3(-1) + \alpha_1 + \alpha_2 + \alpha_3 \\
&= -\alpha_1^2 - \alpha_3^2 - 2\alpha_1\alpha_3 + 2\alpha_3 \\
&= -\frac{1}{4}\alpha_3^2 - \alpha_3^2 - \alpha_3^2 + 2\alpha_3 \\
&= -\frac{9}{4}\alpha_3^2 + 2\alpha_3
\end{aligned}
$$

$$
\begin{aligned}
\frac{\delta}{\delta\alpha_3} &= 0 \\
-\frac{9}{2}\alpha_3 + 2 &= 0 \\
\alpha_3 &= \frac{4}{9}
\end{aligned}
\qquad \Rightarrow \qquad
\begin{aligned}
\alpha_1 &= \frac{2}{9} \\
\alpha_2 &= \frac{2}{9} \\
\alpha_3 &= \frac{4}{9}
\end{aligned}
$$

HOMEWORK 04

Part 2: With the values from part 1, solve for w.



$$
\begin{aligned}
w &= \sum_{i=1}^{3} y_i \alpha_i x_i \\
&= (1)(\frac{2}{9})[0,1] + (1)(\frac{2}{9})[-1,0] + (-1)(\frac{4}{9})[1,-1] \\
&= [0, \frac{2}{9}] + [-\frac{2}{9}, 0] + [-\frac{4}{9}, \frac{4}{9}] \\
&= [-\frac{2}{3}, \frac{2}{3}]
\end{aligned}
$$

HOMEWORK 04

Part 3: Solve for w_0. It can be derived from any support vector with the equation $w_0 = y_i - w^T x_i$. In practice, we average over all support vectors because of numerical imprecision, but in this case you can solve using multiple support vectors to verify that your previous solutions were correct.

$$\begin{aligned} w_0 &= y_i - w^T x_i \\ &= 1 - (-\frac{2}{3}, \frac{2}{3})[0, 1] \\ &= 1 - \frac{2}{3} \\ &= \frac{1}{3} \end{aligned}$$

HOMEWORK 04

Part 4: $w^T(x,y) + w_0 = 0$ describes the decision boundary, with the two components of w projecting into the x and y axes, respectively. Convert that into the familiar line format, $y = mx + b$. Think of this as multiplying w by (x, y) to indicate its components and then isolate y.

$$\begin{aligned} w(x,y) + w_0 &= 0 \\ (-\frac{2}{3}, \frac{2}{3})[x,y] + \frac{1}{3} &= 0 \\ -\frac{2}{3}x + \frac{2}{3}y + \frac{1}{3} &= 0 \\ \frac{2}{3}y &= \frac{2}{3}x - \frac{1}{3} \\ y &= x - \frac{1}{2} \end{aligned}$$

HOMEWORK 04

Part 5: Solve for the size of the margin. This is just $m = \frac{1}{\|w\|}$. Your solution can be verified graphically (see the plot in part 1).

$$
\begin{aligned}
m &= \frac{1}{\|w\|} \\
&= \frac{1}{\sqrt{(}-\frac{2}{3}^2 + \frac{2}{3}^2)} \\
&= \frac{1}{\sqrt{(}\frac{4}{9} + \frac{4}{9})} \\
&= \frac{3}{\sqrt{(}8)}
\end{aligned}
$$