Machine Learning Scibe

Scribed by: Neeraj Saini(20CS60R41), Subham Patidar(20CS60R43)

25,March 2021

1 Introduction

Sometimes in practical situation, we often encounter that we can see observed outcome but we do not see player's internal behaviour. Suppose a robot is moving, we are observed what is output at steps X_1, X_2, X_3, X_4 but we do not see internal states like Y_1 that gives X_1, Y_2 that gives X_2 and so on.

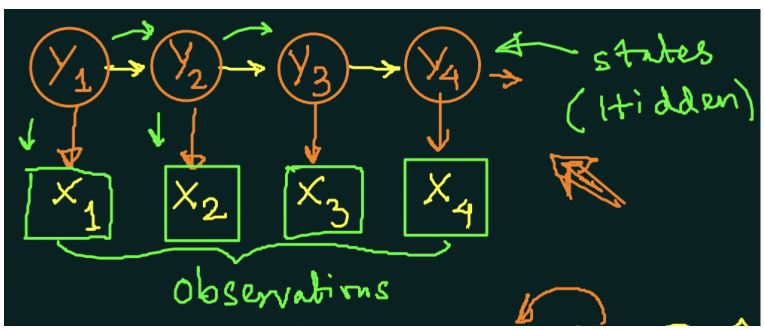

Figure 1: Example of Robot's Hidden Markov Model

Hence we have only clue about the outcomes it produces. i.e in the form of probability depending upon simulation of robot such kind of model is called Hidden Markov Model. Hidden because internal states is not observable only outcome is observable

Lets define the possible distribution of X and Y

Y belongs to Multinominal distribution (discrete)

X belongs to Multinominal distribution if behaviour is discrete or gausian distribution (continuous)

2 Markov Process

consider a sequence of state variables$X_1, X_2, ..., X_t$. A Markov model embodies the Markov assumption on the probabilities of this sequence: that Markov assumption when predicting the future, the past doesn't matter, only the present.

Assumption:

$$P(Y_{t+1}|Y_tY_{t-1}...Y_1) = P(Y_{t+1}|Y_t) \tag{1}$$

Initial Distribution :

$$P(Y_t) : Prior\ (Multinominal\ distribution)$$

Transition Distribution :

$$P(Y_t|Y_{t-1}) : Multinominal\ distribution$$

Emission Distribution :

$$P(X_t|Y_t) : Multinominal\ distribution\ if\ discrete\ or\ Gausian\ distribution\ if\ continous$$

Joint probability Distribution :

$$P(Y_{1..t}X_{1...t}) = P(y_1)\prod_{i=1}^{t-1} P(Y_{i+1}/Y_i)\prod_{i=1}^{t} P(X_i/Y_i) \tag{2}$$

Highlighted Green part can be solved using iterative or recursively

If we know $X_1, X_2, ...X_t$ then we can easily findout Y_t state with the help of HMM, this is called monitoring

Common tasks:

- Monitoring: $P(Y_t|X_{1...t}$
- Prediction: $P(Y_{t+k}|X_{1...t}$
- Hindsight: $P(Y_k|X_{1...t}$ where $k < t$

2.1 Monitoring

$$let\ M^t = P(y_t|X_{1...t}) \tag{3}$$

$$\begin{aligned}
P(y_t|X_{1...t}) &\propto P(Y_t|Y_t, X_{1,2,..t})P(Y_t|X_{1,2,..t}) \quad by\ Bayes\ theorem \\
&= P(X_t|Y_t)P(Y_t|Y_t, X_{1,..t}) \quad By\ \ Conditional\ independence \\
&= P(X_t|Y_t)\sum_{y_{t-1}} P(Y_tY_{t-1}|X_{1,..t}) \quad By\ \ marginalization \\
&= P(X_t|Y_t)\sum_{y_{t-1}} P(Y_t|Y_{t-1}X_{1,..t-1})P(Y_{t-1})|X_{1,..t-1} \quad By\ \ Chain's\ rule \\
&= P(X_t|Y_t)\sum_{y_{t-1}} P(Y_{t-1})|X_{1,..t-1}\ M^{t-1} \\
&\quad Higlighted\ green\ part\ can\ be\ solved\ through\ recursively
\end{aligned} \tag{4}$$

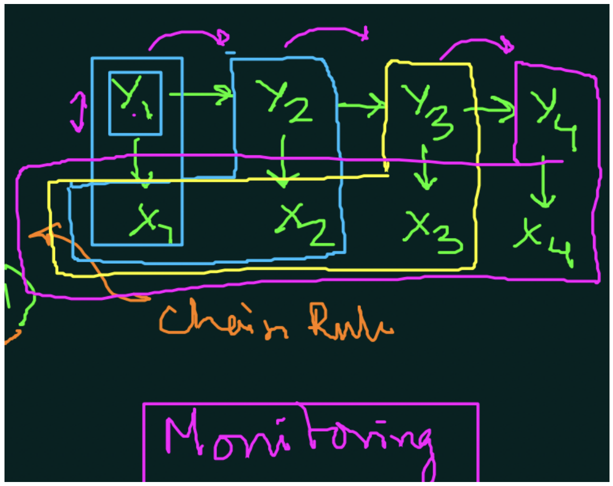

Figure 2: Monitoring(Forward Algorithm)

2.2 Prediction

We can easily see that prediction helps in agents to predict the weather whether it is sunny or rainy or dry.

$$
\begin{aligned}
& P^k \, P(y_{t+k}|X_{1...t}) \qquad where\ k > t \\
&= \sum_{Y_{t+k-1}} P(Y_{t+k}Y_{t+k-1}|X_{1,..t}) \quad By\ \ marginalization \\
&= \sum_{Y_{t+k-1}} P(Y_{t+k}|Y_{t+k-1}, X_{1,..t}) \quad By\ \ chain'srule \\
&= \sum_{Y_{t+k-1}} P(Y_{t+k}|Y_{t+k-1}) \, P^{k-1} \\
& higlighted\ gren\ part\ can\ be\ solved\ through\ recursively
\end{aligned}
\tag{5}
$$

let's try to understand how this recursion proceed with our observation.

Now we want to predict Y_4 and we only have the view of X_1 and X_2 . First

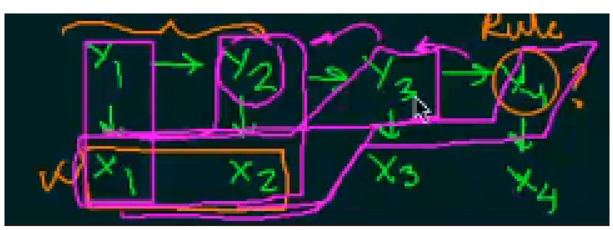

Figure 3: prediction(forward Algorithm)

we try to get Y_2 by using monitoring algorithm and then recursive algorithm of prediction helps to predict Y_4 .

2.3 Hindsight

Sometimes it is required in speech recognition. Simply it connects the dots between the speech.

$$
\begin{aligned}
&P(y_k|X_{1...t}) \qquad where\ k < t \\
&= P(Y_k X_{k+1...t}|X_{1,..k}) \quad By\ conditionalindependence \\
&= P(Y_k|X_{1...k})p(X_{k+1,..t}|Y_k) \\
&(P(Y_k|X_{1...k}) = M^k and \\
&(p(X_{k+1,..t}|Y_k) = H^k \quad By\ chain's rule \\
&p(X_{k+1,..t}|Y_k \\
&= \sum_{Y_{k+1}} P(Y_{k+1}, X_{k+1...t}|Y_k) \quad By\ \ marginalization \\
&= \sum_{Y_{k+1}} P(Y_{k+1}|Y_k).P(X_{k+1}|Y_{k+1}).P(X_{k+2...t}|Y_{k+1})[\ H^{k+1}]
\end{aligned} \tag{6}
$$

from the above method we have an interesting observation. we will be

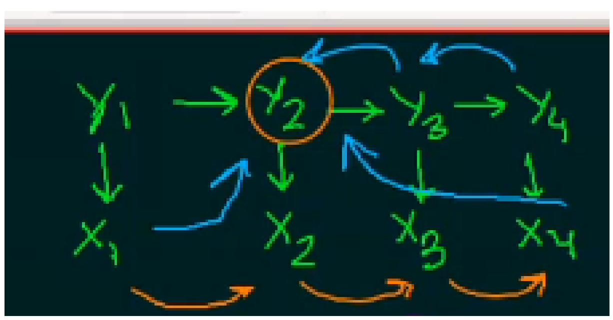

Figure 4: hindsight(forward backward Algorithm)

dictate Y_4 state because X_4 observation is related to Y_4 .
So to get at Y_2 not only forward pass from X_1 matters but also backward steps from $X_4 X_3 X_2$ also matters.
Therefore the above mention algorithm of hindsight is also called forward-backward algorithm.

suppose we go through the series of sequence of states $Y_1 -> Y_2 -> Y_3$ but we do not know this series because they are hidden, we only knows $X_1 X_2 X_3$. so we want that what is the most likely explanation of Y passing through this sequences.

$$max_{Y_{1...t}} P(y_{1...t}|X_{1...t}) \qquad most\ likely\ explanation \tag{7}$$

2.4 Viterbi Algorithm

In this algorithm we try to make a likely explanation.

In this algorithm we try to make a most likely explanation, whether we do

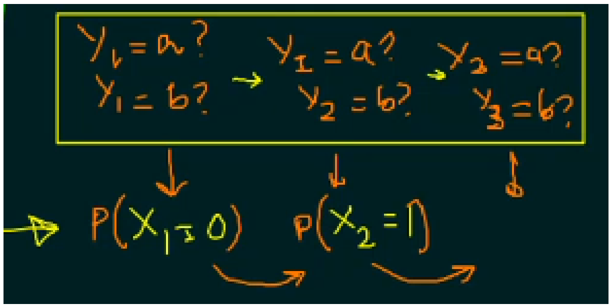

Figure 5: viterbi Algorithm)

not know any information about $Y_1 = a?, Y_1 = b?, Y_2 = a?, Y_2 = b?$. We only knows $P(X_1 = 0, P(X_2 = 1$. So which is the most likely sequences governed to this situation $X_1 = 0, X_2 = 1$.

To maximize which one gives higher probability we use the dynamic programming algorithm.

$$\begin{aligned} &Y^{max}_{1...t}(P(y_{1...t}|X_{1...t}) \\ &= Y^{max}_t(P(X_t|Y_t)Y^{max}_{1..t-1}(P(Y_{1..t-1}|X_{1..t}) \qquad Dynamic\ Programming \end{aligned} \tag{8}$$

In the real life scenario if we face some problem over hidden markov model (HMM) like: if the observations are partial as given below:

$$\begin{aligned} &P(y) - > partial \\ &P(y_{t)}|X_t) - > partial \end{aligned} \tag{9}$$

Then we have to go over expectation maximization algorithm which can be also applied over HMM.

Contributions:
Neeraj Saini: Introduction, Markov process, Monitoring
Shubham Patidar : Prediction, Hindsight, Veterbi algorithm