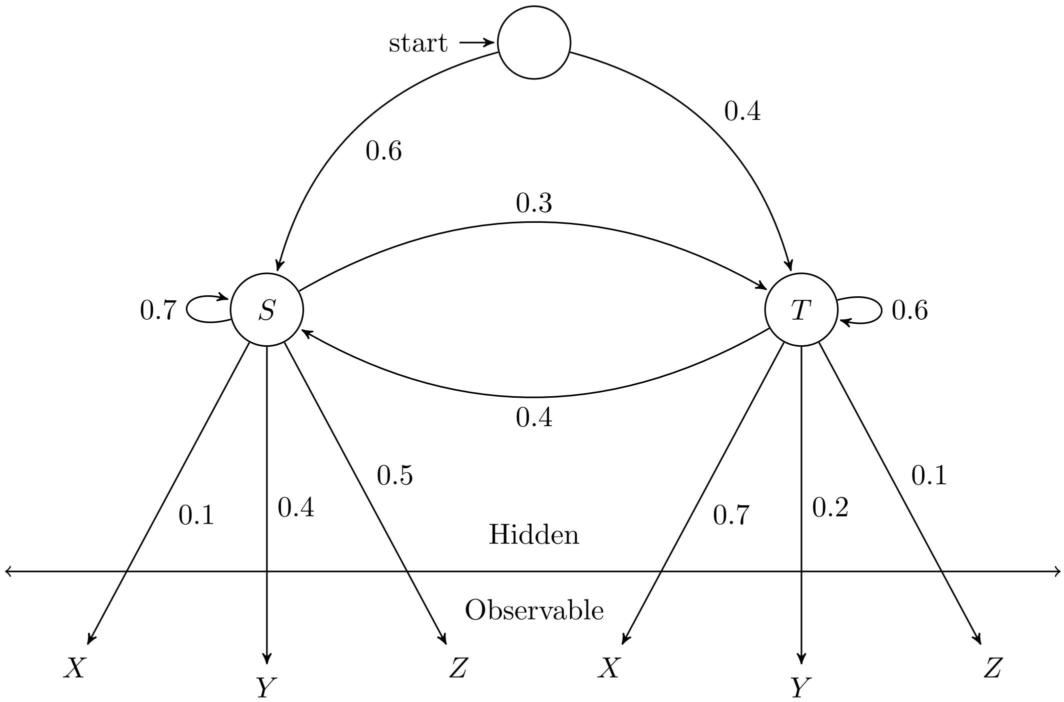
HIDDEN MARKOV MODELS

Hidden Markov Models(HMMs) are statistical tools to model sequential observations with the assumption that states of the system generating them follow a Markov process but these states are unobservable/hidden. However, an observation at any point is related to the underlying hidden state the system is in at that point.

1 Example



In this diagram, nodes $\{S, T\}$ represent the hidden states and $\{X, Y, Z\}$ represent observable states. The unlabelled node is the start node. These states together with the associated probabilities fully characterize the HMM. The state transition probabilites denoted by A can be represented as,

$$A = \begin{array}{c c} & \begin{array}{cc} S & T \end{array} \\ \begin{array}{c} S \\ T \end{array} & \begin{bmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{bmatrix} \end{array}$$

The emmission probabilities are denoted by B.

$$B = \begin{array}{c c} & \begin{array}{ccc} X & Y & Z \end{array} \\ \begin{array}{c} S \\ T \end{array} & \begin{bmatrix} 0.1 & 0.4 & 0.5 \\ 0.7 & 0.2 & 0.1 \end{bmatrix} \end{array}$$

Lastly, initial state distribution is denoted by π.

$$\pi = \begin{array}{c} \begin{array}{cc} S & T \end{array} \\ \begin{bmatrix} 0.6 & 0.4 \end{bmatrix} \end{array}$$

2 The three problems of HMM

2.1 The Evaluation Problem

Given a HMM, $\lambda = (A, B, \pi)$ and a set of observations O, find $P(O|\lambda)$ (probability that the observations were generated by the model).

Consider the HMM in the example above and let $O = (ZXY)$.

$$\begin{aligned} P(ZXY|\lambda) = {} & P(SSS, ZXY|\lambda) \\ & + \ P(SST, ZXY|\lambda) \\ & + \ P(STS, ZXY|\lambda) \\ & + \ P(STT, ZXY|\lambda) \\ & + \ P(TSS, ZXY|\lambda) \\ & + \ P(TST, ZXY|\lambda) \\ & + \ P(TTS, ZXY|\lambda) \\ & + \ P(TTT, ZXY|\lambda) \end{aligned}$$

Each of the terms on R.H.S. can be calculated using the following law of probability.

$$\begin{aligned} P(a, b|c) &= \frac{P(a, b, c)}{P(c)} \\ &= \frac{P(a, c)}{P(c)} * \frac{P(a, b, c)}{P(a, c)} \\ &= P(a|c) * P(b|a, c) \end{aligned}$$

Table 1: $P(ZXY|\lambda)$ calculation

HS	OS	$P(HS\|\lambda)$	$P(OS\|HS,\lambda)$	$P(HS,OS\|\lambda)$
SSS	ZXY	0.6*0.7*0.7=0.294	0.5*0.1*0.4=0.020	0.005880
SST	ZXY	0.6*0.7*0.3=0.126	0.5*0.1*0.2=0.010	0.001260
STS	ZXY	0.6*0.3*0.4=0.072	0.5*0.7*0.4=0.140	0.010080
STT	ZXY	0.6*0.3*0.6=0.108	0.5*0.7*0.2=0.070	0.007560
TSS	ZXY	0.4*0.4*0.7=0.112	0.1*0.1*0.4=0.004	0.000448
TST	ZXY	0.4*0.4*0.3=0.048	0.1*0.1*0.2=0.002	0.000096
TTS	ZXY	0.4*0.6*0.4=0.096	0.1*0.7*0.4=0.028	0.002688
TTT	ZXY	0.4*0.6*0.6=0.144	0.1*0.7*0.2=0.014	0.002016
			$P(ZXY\|\lambda)$	0.030028

Calculation in this manner is expensive because probabilities corresponding to all permutations of hidden states of the same length as the length of observed sequence have to be calculated.

2.2 The Decoding Problem

2.3 The Learning Problem