Lecture 11: Learning from Hidden Markov Model

Lokesh Kumar

March 5, 2008

1 Background

As discussed in the Lecture 10, we can uniquely define a Hidden Markov Model (HMM) by specifying

$$M = (Q,\ \Sigma,\ A,\ e)$$

where:

Q = The set of all states that the model can be in

Σ = The set of all alphabets that each state can be in.

$$A = \text{state transition probability} = \begin{pmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & a_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

e = emission probability $= e_{Q_i}(j)\ \forall j : j \ \epsilon \ \Sigma$

Going back to our favorite example from previous lecture, we can model a casino dealer's operations using an HMM by specifying the above parameters as:

$$Q = \{Fair, Loaded\} \tag{1.1}$$

$$\Sigma = \{1, 2, 3, 4, 5, 6\} \tag{1.2}$$

$$A = \begin{pmatrix} a_{F,F} & a_{F,L} \\ a_{L,F} & a_{L,L} \end{pmatrix} \tag{1.3}$$

$$e_{fair}(i) = \frac{1}{6}, i = 1, 2, \ldots, 6 \tag{1.4}$$

$$e_{loaded}(i) = \begin{cases} \frac{1}{2} & \text{i = 6} \\ \frac{1}{10} & \text{i = 1, 2, \ldots, 5} \end{cases} \tag{1.5}$$

2 Inferring parameters

So far we have used the HMM to infer

1. **Decoding Problem:** Maximum likelihood of states, given a sequence of observations of values of die

2. **Inference Problem:** Probability of observing any sequence of values of die

Another question that we can ask is - given a sequence of observation of values of die, what are the most likely values of the HMM parameters. To put it more precisely -

Learning Problem
Given a sequence of observations X_1, X_2, X_3, ..., X_n, what are the values of transition probabilities A and emission probabilities $e_{F/L}$ that best explain these observations ?

2.1 ML Approach

If we also know the sequence of states $Z_1 \to Z_2 \to Z_3 \to \ldots \to Z_n$, we can solve this problem by counting the number of state transitions.

$$\begin{aligned} N_{FF} &\doteq \text{Number of transitions from Fair to Fair} && (2.1) \\ N_{FL} &\doteq \text{Number of transitions from Fair to Loaded} && (2.2) \\ N_{LF} &\doteq \text{Number of transitions from Loaded to Fair} && (2.3) \\ N_{LL} &\doteq \text{Number of transitions from Loaded to Loaded} && (2.4) \end{aligned}$$

Therefore,

$$\tilde{a_{FF}} = \frac{N_{FF}}{N_{FF} + N_{FL}} \tag{2.5}$$

$$\tilde{a_{FL}} = \frac{N_{FL}}{N_{FF} + N_{FL}} \tag{2.6}$$

$$\tilde{a_{LF}} = \frac{N_{LF}}{N_{LF} + N_{LL}} \tag{2.7}$$

$$\tilde{a_{LL}} = \frac{N_{LL}}{N_{LF} + N_{LL}} \tag{2.8}$$

We can similary count the number of times the dice gives a particular value i in each state. Solving it for the emission probabilities in Fair state, we have

$$e_{fair}(i) = \frac{N_{i,F}}{N_F} \ \forall \text{ i } \epsilon \ \{ \text{ 1,2,} \ldots \text{ , 6 } \} \tag{2.9}$$

where

$$N_{i,F} \doteq \text{Number of times observed value was i in Fair state} \tag{2.10}$$

$$N_F \doteq \text{Number of times model is in Fair state} \tag{2.11}$$

Similary, we can find the emission probabilities e_L. The above approach of finding the parameter values is called the Maximum Likelihood method since it gives the best way of fitting HMM's parameters to the given data.

2.2 EM Approach

The method described above works well when we know the sequence of states of HMM. However, this may not always be possible since the states are typically *hidden* in HMM. In such scenarios we resort to using iterative approaches for the problem. The basic idea is similar to Expectation-Maximization (EM) algorithm where we start with an initial estimate of forward probabilities and backward probabilities and iteratively

- Calculating the transition and emission probabilities from forward and backward probabilities
- Using the probabilities computed in the previous step to update the forward and backward probabilities

This iteration is repeated till it converges.

More specifically, let us define the transition probability from State i to j for the t^{th} observation as

$$\xi_t(i,j) = P(Z_t = i, Z_{t+1} = j | X_1, X_2, \ldots, X_n, A, e) \tag{2.12}$$

$$\Rightarrow \xi_t(i,j) = \frac{P\{Z_t = i, Z_{t+1} = j, X_1, X_2, \ldots, X_n | A, e\}}{P\{X_1, X_2, \ldots, X_n | A, e\}} \tag{2.13}$$

This can also be expressed using the forward and backward probabilities as

$$\xi_t(i,j) = \frac{f_t(i) a_{ij} b_{t+1}(j) e_j(X_{t+1})}{\sum_{i=1}^{n} \sum_{j=1}^{n} f_t(i) a_{ij} b_{t+1}(j) e_j(X_{t+1})} \tag{2.14}$$

The posterior probability of model to be in State i during t^{th} observation can thus be expressed as

$$\gamma_t(i) = \sum_{j=1}^{n} \xi_t(i,j) \tag{2.15}$$

The probabilities defined in equations 2.14 and 2.15 are calculated in the Expectation step using an intial or previous iteration estimate of transition probabilities A and emission probabilities e. These probabilities are then used in the Maximization step to update A and e using the following identities:

$$\overline{a_{ij}} = \frac{\sum_{t=1}^{L-1} \xi_t(i,j)}{\sum_{t=1}^{L-1} \gamma_t(i)} \tag{2.16}$$

and

$$\overline{e_j(k)} = \frac{\sum_{t=1, X_t=k}^{L} \gamma_t(j)}{\sum_{t=1}^{L} \gamma_t(j)} \tag{2.17}$$

This iteration is continued till update in the values of A and e drops below some predefined convergence threshold. The above approach of using generalize HMM for learning is also called Baum-Welch Algorithm.

3 References

1. L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains", Ann. Math. Statist., vol. 41, no. 1, pp. 164–171, 1970.

2. Hidden Markov Models and the Baum-Welch Algorithm, IEEE Information Theory Society Newsletter, Dec. 2003.

3. Scribe Notes for Lecture - 10, CS 284A, UCI.