Machine Learning-Homework 5

Written Problems

1. (a) Factorized probability distribution for HMM can be written as

$$p(X,Y) = p(x_1)\prod_{t=1}^{T-1} p(x_{t+1}|x_t)\prod_{t'=1}^{T} p(y_{t'}|x_{t'}) \tag{1}$$

(b)

$$p(x_1,y_1) = p(x_1)p(y_1|x_1) \tag{2}$$

(c) From (2)

$$p(x_2,y_1,y_2) = \sum_{x_1} p(x_1,y_1)p(x_2|x_1)p(y_2|x_2) \tag{3}$$

$$= \sum_{x_1} \alpha_1(x_1)p(x_2|x_1)p(y_2|x_2) \tag{4}$$

(d) Generalizing,

$$p(x_t,y_1,\ldots,y_t) = \sum_{x_{t-1}} \alpha(x_{t-1})p(x_t|x_{t-1})p(y_t|x_t) \tag{5}$$

(e)

$$p(y_T|x_{T-1}) = \beta_{T-1}(x_{T-1}) = \sum_{x_T} p(x_T|x_{T-1})p(y_T|x_T)\beta_T(x_T) \tag{6}$$

$$= \sum_{x_T} p(x_T|x_{T-1})p(y_T|x_T) \tag{7}$$

since β_T is assumed to be 1

(f)

$$\beta_{T-2}(x_{T-2}) = \sum_{x_{T-1}} p(x_{T-1}|x_{T-2})p(y_{T-1},y_T|x_{T-1}) \tag{8}$$

$$= \sum_{x_{T-1}} p(x_{T-1}|x_{T-2})p(y_{T-1}|x_{T-1})p(y_T|x_{T-1}) \tag{9}$$

$$= \sum_{x_{T-1}} p(x_{T-1}|x_{T-2})p(y_{T-1}|x_{T-1})\beta_{T-1}(x_{T-1}) \tag{10}$$

Generalizing,

$$\beta_{t-1}(x_{t-1}) = p(y_t,\ldots,y_T|x_t) \tag{11}$$

$$= \sum p(x_t|x_{t-1})p(y_t|x_t)p(y_{t+1},\ldots,y_T|x_t) \tag{12}$$

$$= \sum p(x_t|x_{t-1})p(y_t|x_t)\beta_t(x_t) \tag{13}$$

(g)

$$p(x_t, y_1, \ldots, y_T) = p(x_t, Y) = p(x_t, y_1, \ldots, y_t)p(y_{t+1}, \ldots, y_T|x_t) \tag{14}$$

$$p(x_t, Y) = \alpha_t(x_t)\beta_t(x_t) \tag{15}$$

$$p(x_t|Y) = p(x_t, Y)/p(Y) \tag{16}$$

$$p(x_t|Y) = \frac{\alpha_t(x_t)\beta_t(x_t)}{\sum_{x_t'} \alpha_t(x_t')\beta_t(x_t')} \tag{17}$$

(h) Let,for the first time instant

$$c_1 = \frac{1}{\sum_{x_t} \alpha_1(x_t)} \tag{18}$$

$$\hat{\alpha_1}(x_t) = c_1\alpha_1(x_t) \tag{19}$$

Where N is the number if states and c_t represents normalization at each step. We have $\hat{\alpha_1}(x_t) = c_1\alpha_1(x_t)$
Let us assume that for time instant t

$$\hat{\alpha_t}(x_t) = c_1c_2 \ldots c_t\alpha_t(x_t) \tag{20}$$

Then for time $t+1$ From (5),

$$\hat{\alpha}_{t+1}(x_{t+1}) = c_{t+1} \sum_{x_t} \hat{\alpha_t}(x_t)p(x_{t+1}|x_t)p(y_{t+1}|x_{t+1}) \tag{21}$$

From (20),

$$\hat{\alpha}_{t+1}(x_t) = c_1c_2 \ldots c_tc_{t+1} \sum_{x_t} \alpha_t(x_t)p(x_{t+1}|x_t)p(y_{t+1}|x_{t+1}) \tag{22}$$

$$\hat{\alpha}_{t+1}(x_t) = c_1c_2 \ldots c_tc_{t+1}\alpha_{t+1}(x_t) \tag{23}$$

Thus by induction,(20) holds for all t. Rewriting equation(20),

$$\hat{\alpha_t}(x_t) = \left(\prod_{t'=1}^{t} c_{t'} \right) \alpha_t(x_t) \tag{24}$$

$$\hat{\alpha_t}(x_t) = \mathbf{C_t}\alpha_t(x_t) \tag{25}$$

where $\mathbf{C_t} = \prod_{t'=1}^{t} c_{t'}$

Similarly, doing it for β,
We have $\hat{\beta}_T(x_t) = z_T \beta_T(x_t)$, where $z_T = 1/\sum_{x_t} \beta_T(x_t)$
Let us assume that for time instant t

$$\hat{\beta}_t(x_t) = z_t z_{t+1} \dots z_T \beta_t(x_t) \tag{26}$$

$$\hat{\beta}_{t-1}(x_t) = z_{t-1} \sum_{x_t} p(x_t|x_{t-1}) p(y_t|x_t) \hat{\beta}_t(x_t) \tag{27}$$

From (25),

$$\hat{\beta}_{t-1}(i) = z_{t-1} z_t z_{t+1} \dots z_T \sum_{x_t} p(x_t|x_{t-1}) p(y_t|x_t) \beta_t(x_t) \tag{28}$$

$$\hat{\beta}_{t-1}(x_t) = z_{t-1} z_t z_{t+1} \dots z_T \beta_{t-1}(x_t) \tag{29}$$

By induction, (26) is true for all time instants.
We have,

$$\hat{\beta}_t(x_t) = \left(\prod_{t'=t}^{T} z_{t'} \right) \beta_t(x_t) \tag{30}$$

$$\hat{\beta}_t(x_t) = \mathbf{Z_t} \beta_t(x_t) \tag{31}$$

where $\mathbf{Z_t} = \left(\prod_{t'=t}^{T} z_{t'} \right)$
The update rule for Baum Welsh is

$$p(x_{t'+1}|x_{t'}) \leftarrow \frac{\sum_{t=1}^{T-1} p(x_{t+1}, x_t|Y)}{\sum_{t=1}^{T} p(x_t|Y)} \tag{32}$$

$$p(x_{t'+1}|x_{t'}) \leftarrow \frac{\sum_{t=1}^{T-1} \alpha_t(x_t) p(x_{t+1}|x_t) \beta_{t+1}(x_{t+1}) p(y_{t+1}|x_{t+1})}{\sum_{t=1}^{T} \alpha_t(x_t) \beta_t(x_t)} \tag{33}$$

$$p(x_{t'+1}|x_{t'}) \leftarrow \frac{\sum_{t=1}^{T-1} \frac{\hat{\alpha}_t(x_t)}{\mathbf{C_t}} p(x_{t+1}|x_t) \frac{\hat{\beta}_{t+1}(x_{t+1})}{\mathbf{Z_{t+1}}} p(y_{t+1}|x_{t+1})}{\sum_{t=1}^{T} \frac{\hat{\alpha}_t(x_t)}{\mathbf{C_t}} \frac{\hat{\beta}_{t+1}(x_{t+1})}{\mathbf{Z_t}}} \tag{34}$$

We have that the scaling factors are same for both α, β i.e $c_t = z_t$ Thus we have $\mathbf{C_t Z_{t+1}} = \prod_{t=1}^{T} c_t$ and $\mathbf{C_t Z_t} = c_t \prod_{t=1}^{T} c_t$
Thus,

$$p(x_{t'+1}|x_{t'}) \leftarrow \frac{\left(\sum_{t=1}^{T-1} \alpha_t(x_t) p(x_{t+1}|x_t) \beta_{t+1}(x_{t+1}) p(y_{t+1}|x_{t+1}) \right) / \mathbf{C_t Z_{t+1}}}{\left(\sum_{t=1}^{T} \alpha_t(x_t) \beta_t(x_t)/c_t \right) / \mathbf{C_t}.\mathbf{Z_{t+1}}} \tag{35}$$

$$p(x_{t'+1}|x_{t'}) \leftarrow \frac{\sum_{t=1}^{T-1} \alpha_t(x_t) p(x_{t+1}|x_t) \beta_{t+1}(x_{t+1}) p(y_{t+1}|x_{t+1})}{\sum_{t=1}^{T} \alpha_t(x_t) \beta_t(x_t)/c_t} \tag{36}$$

Where c_t is the placeholder factor Z_t mentioned in the question
Thus the update equation remains the same, hence we can normalize at each step.

Programming Assignment Writeup

Sparsification drastically reduced the execution time for markovChainTrain.m where the transition probabilities and priors were calculated.HMM's perform well while producing the output text even though we are considering only the bigrams(transition from t to $t+1$).Quite a lot of sentences make sense.For learning the Gaussians, HmmInferStates.m and myHmmTrain.m was modified where the model means and sigma were updated at every iteration by treating the hidden variable in the HMM as the latent variable. RunSyntheticExperiments was run and the results are in "html" folder.The roboObama.txt contains output of the runObama.m