Machine Learning 2

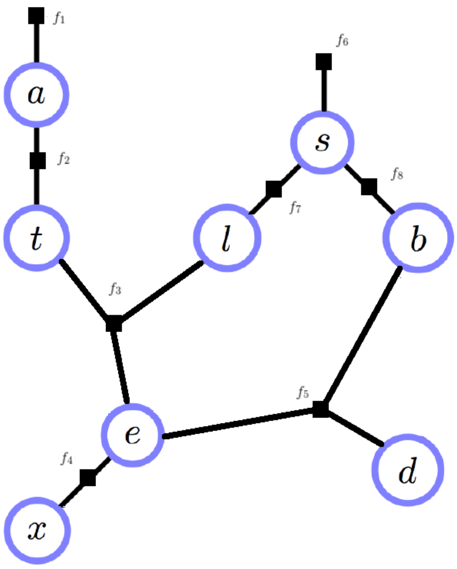
Homework 3 – Group 8

Anton Zeltser

Harsha Nandeeshappa

Problem1. Factor Graphs and Messages

1)



2)

$$f_5 = f$$

$$\mu_{f\to d}(d) = \sum_{e,b} f(e,b,d)\mu_{e\to f}(e)\mu_{b\to f}(b)$$

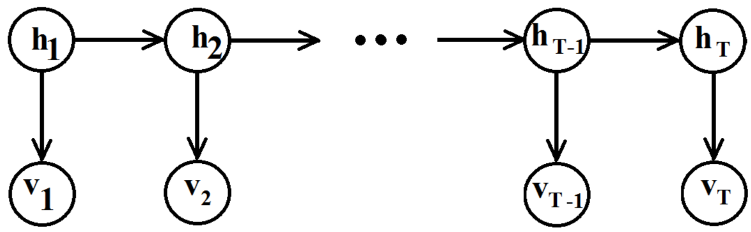
3)

$$\mu_{e\to f}(e) = \mu_{f_3\to e}(e)\mu_{f_4\to e}(e)$$

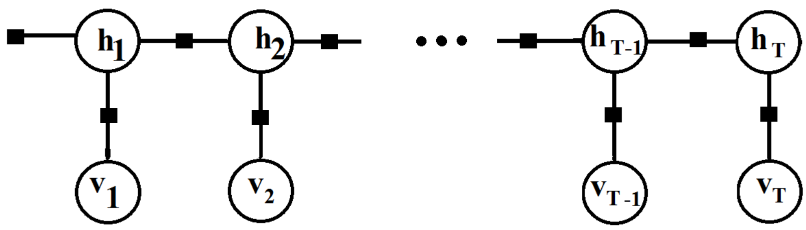
$$\mu_{b\to f}(e) = \mu_{f_8\to e}(e)$$

Problem2. Hidden Markov Model

1)



2)



3)

$$p\left(h_t \middle| v_1,\dots,v_T\right)=\frac{p\left(h_t,v_1\dots,v_T\right)}{p\left(v_1\dots,v_T\right)}\propto p\left(h_t,v_1\dots,v_T\right)$$

$$p\left(h_t,v_1\dots,v_T\right)=\sum_{h_1,\dots h_T,except\ h_t} p\left(h_1,h_2..h_T,v_1\dots,v_T\right)=$$

$$\sum_{h_1,\dots h_T,except\ h_t}\varphi\left(h_1\right)\varphi\left(h_1,h_2\right)\varphi\left(v_1,h_1\right)\varphi\left(h_2,h_3\right)\varphi\left(v_2,h_2\right)\cdots\ \varphi\left(h_T,h_{T-1}\right)\varphi\left(v_T,h_T\right)=$$

$$\sum_{h_1,\dots h_{t-1}}\varphi\left(h_1\right)\varphi\left(v_1,h_1\right)\prod_{t=2}^{T}\varphi\left(h_i,h_{i-1}\right)\varphi\left(v_i,h_i\right)\sum_{h_{t+1},\dots h_T}\varphi\left(h_1\right)\varphi\left(v_1,h_1\right)\prod_{t=2}^{T}\varphi\left(h_i,h_{i-1}\right)\varphi\left(v_i,h_i\right)=$$

$$\mu_{f_{left}\to h_t}\left(h_t\right)\cdot\mu_{f_{right}\to h_t}\left(h_t\right)$$

4)

$$p\left(h_t,h_{t+1} \middle| v_1,\dots,v_T\right)=\frac{p\left(h_t,h_{t+1},v_1\dots,v_T\right)}{p\left(v_1\dots,v_T\right)}\propto p\left(h_t,h_{t+1},v_1\dots,v_T\right)$$

$$p\left(h_t,h_{t+1},v_1\dots,v_T\right)=\sum_{h_1,\dots h_T,except\ h_t and\quad h_{t+1}} p\left(h_1,h_2..h_T,v_1\dots,v_T\right)=$$

$$\sum_{h_1,\dots h_T,except\ h_t\ and\ h_{t+1}}\varphi\left(h_1\right)\varphi\left(h_1,h_2\right)\varphi\left(v_1,h_1\right)\varphi\left(h_2,h_3\right)\varphi\left(v_2,h_2\right)\cdots\ \varphi\left(h_T,h_{T-1}\right)\varphi\left(v_T,h_T\right)=$$

$$\sum_{h_1,\dots h_{t-1}}\varphi\left(h_1\right)\varphi\left(v_1,h_1\right)\prod_{t=2}^{T}\varphi\left(h_i,h_{i-1}\right)\varphi\left(v_i,h_i\right)\sum_{h_{t+2},\dots h_T}\varphi\left(h_1\right)\varphi\left(v_1,h_1\right)\prod_{i=2}^{T}\varphi\left(h_i,h_{i-1}\right)\varphi\left(v_i,h_i\right)$$

Problem 3.

1 -- v_t *and* h_t corresponds to visible and hidden state respectively.

v_t is the visible(observed) states i.e the keypad buttons that are pressed. It's domain is 2,3,4,5,6,7,8,9 .

h_t corresponds to hidden states i.e actual alphabet the user intended to type while pressing a keypad button. It could be any alphabet depending on the language used. We have chooses the language is English. It's domain is a,b,c,d,e,f,g,h,i,j,k,l,m,n,o,p,q,r,s,t,u,v,w,x,y,z.

2 – We need to learn transition matrix(P($h_t|h_{t+1}$)) , Initial matrix((P(h_t))),

Emmision Matrix ((P(v|h_t)) for t = 1 to T

3 – Transition matrix – using corpus/book

Initial Matrix – Using book or text messages

Emmision Matrix – By observing the behavior of keypad when key is pressed. i.e what letters are output when a key is pressed.

4 – Inference step is finding h_t given v_t.

How quickly – In this experiment, it depends on the n-grams we chose. It can be done quickly if we use log of the terms which replaces multiplication by additions.

5 – We think HMM is good model for this problem.

Also it depends on how many states are there in h_i and that depends on the n-gram we choose.

We can infer better (meaning more accurate) with bigger n in n-grams, but this takes more computations.

Problem 4

1 –

Sum of the rows of P =

[2.15606161 0.14042886 0.46731331 0.50520068 6.07112797 0.34045886
 0.29531875 0.97819872 2.4403277   0.01834861 0.05808857 0.90730752
 0.29239429 1.28481483 2.57223902 0.49137424 0.01654401 1.42908813
 1.31282881 1.39167035 1.93624418 0.25365599 0.07372152 0.08910132
 0.42961227 0.0485299]

Sum of the rows of Q =

[0.33333333 0.33333333 0.33333333 0.33333333 0.33333333 0.33333333
 0.33333333 0.33333333 0.33333333 0.33333333 0.33333333 0.33333333
 0.33333333 0.33333333 0.33333333 0.25   0.25   0.25   0.25
 0.33333333 0.33333333 0.33333333 0.25   0.25   0.25   0.25   ]

Sum of the rows of R =

[1.]

Sum over the columns of P =

[1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.
 1. 1. 1. 1. 1. 1. 1. 1.]

Sum over the columns of Q =

[1. 1. 1. 1. 1. 1. 1. 1.]

Sum over the columns of R =

[0.11175651 0.03844518 0.0436817   0.04429776 0.02921892 0.04639531
 0.01348001 0.05792446 0.0670187   0.00610194 0.00418042 0.02559589
 0.03869454 0.02187019 0.07309131 0.05152915 0.00457646 0.02757609
 0.05349468 0.14958563 0.01122112 0.01029703 0.06160616 0.00164283
 0.00605794 0.00066007]

2.

P – Transition Matrix

Q- Emmision Matrix

R- Initial Matrix

3. <code>

4.<code>

5. – We need to compute

$T[0] = R[v_1]$

Then compute

$T[v_i]$ $*Q[v_i ,h_{i+1}]$ $* P[h, h_{i+1}]$

And update T with maximum likelihood every iteration.

6<code>

7

Given sequence - 6224463

Most likely h - oachind

Given sequence - 53276464

Most likely h - learogng