Machine Learning II – Assignment 3

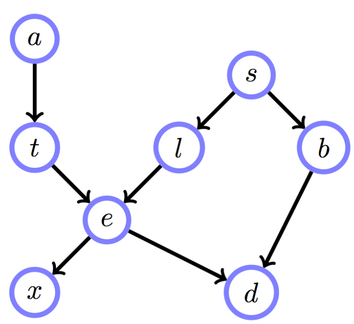
Stefan Roth, Tobias Plötz, Uwe Schmidt

November 26, 2014

This homework is due December 10, 2014 at 20:00.
Please read the instructions carefully!

Important: For task 4 in this assignment, you may use a programming language of your choice. In addition to submitting your solution in Moodle as usual, you will have to demonstrate and explain the code in person in the week before the Winter break (December 15–19); details about this will be announced via Moodle. If your program does not work on the machines in our computer pool, you will have to use your own computer for demonstration. You may not work any further on your code after the submission deadline.

1 Factor Graphs and Messages (4 Points)



1. Draw a factor graph representation for this network
2. Let f be the factor associated with d. Write out the message $\mu_{f\to d}(d)$.
3. Write out the messages $\mu_{e\to f}(e)$ and $\mu_{b\to f}(b)$.

2 Hidden Markov Model (6 Points)

Consider the hidden Markov Model (HMM)

$$p(v_1, \ldots, v_T, h_1, \ldots, h_T) = p(h_1)p(v_1 \mid h_1) \prod_{t=2}^{T} p(v_t \mid h_t)p(h_t \mid h_{t-1}) \quad (1)$$

for which each h_t can take states in $\{1, \ldots, H\}$ and v_t in $\{1, \ldots, , V\}$ for all $t = 1, \ldots, T$.

1. Draw a Belief network representation of the above distribution
2. Draw a Factor Graph representation of the above distribution
3. Use the Factor Graph to derive a Sum-Product algorithm to compute marginals $p(h_t \mid v_1, \ldots, v_T)$. Explain the sequence order of messages passed on your Factor Graph
4. Explain how to compute $p(h_t, h_{t+1} \mid v_1, \ldots, v_T)$

3 Hidden Markov Model for T9 (5 Points)

T9 is a text input method to increase the usability for text entry using keypads with limited number of keys (e.g. keys 0-9 on cell phones). See your old cell phone or the Internet for more details. For this we will use the hidden Markov Model (HMM) model

$$p(\mathbf{v}, \mathbf{h}) = p(v_1, \ldots, v_T, h_1, \ldots, h_T) = p(h_1)\, p(v_1|h_1) \prod_{t=2}^{T} p(v_t|h_t)\, p(h_t|h_{t-1}) \quad (2)$$

as described in the previous task. To this end v_t are used to model the variables that are observed (the "visible" variables) and h_t are the variables we want to infer (the "hidden" variables).

1. What are the v_t, h_t in this case? What is their domain?
2. What are the parameters we need to learn?
3. Where will we get training data from?
4. What is the inference step? Can it be done quickly?
5. Is the HMM a good model for this problem, what are possible extensions and what are the problems that may come with them?

4 T9 continued (15 Points)

In the HMM, each h_t can take states in $\{1, \ldots, H\}$ and v_t in $\{1, \ldots, V\}$ for all $t = 1, \ldots, T$. Let us store the CPT for the transitions between two hidden states using the matrix $\mathbf{P}$. That is

$$p(h_t = i|h_{t-1} = j) = P_{ij} \quad (3)$$

where the transitions do not depend on the time t (are stationary). Similarly

$$p(v_t = i | h_t = j) = Q_{ij} \tag{4}$$

and

$$p(h_1 = j) = R_j. \tag{5}$$

Note that $\mathbf{P}$ is a $H \times H$ matrix whereas $\mathbf{Q}$ is of size $V \times H$ and $\mathbf{R}$ an $H \times 1$ vector. Assume we have observed training data pairs for both visible and hidden variables

$$\mathcal{D} = \{(\mathbf{h}^1, \mathbf{v}^1), (\mathbf{h}^2, \mathbf{v}^2), \ldots, (\mathbf{h}^N, \mathbf{v}^N)\}. \tag{6}$$

1. What is the sum of the rows of $\mathbf{P}, \mathbf{Q}, \mathbf{R}$? What is the sum over the columns of $\mathbf{P}, \mathbf{Q}, \mathbf{R}$?

2. We want to train the HMM using $\mathcal{D}$ by maximum likelihood. What are the parameters $\mathbf{P}, \mathbf{Q}, \mathbf{R}$?

3. Download a suitable large text file of your choice (e.g. a book). Write a function that counts how often two characters follow each other (e.g. *aa*, *ab*, *ac*, …; as in *aa*chen, r*ab*iat, aufw*ac*hen, …). This can be stored in a count matrix $\mathbf{N}$ of size 26×26. Leave out punctuation and do not count transitions over whitespaces.

4. Set $\mathbf{P}, \mathbf{Q}, \mathbf{R}$ to their maximum likelihood estimates (with the help of $\mathbf{N}$).

5. Given $\mathbf{P}, \mathbf{Q}, \mathbf{R}$, assume we are presented a new visible variable $\mathbf{v} = (v_1, \ldots, v_T)$. Now we want to infer the most likely configuration $\mathbf{h} = (h_1, \ldots, h_T)$. What are the messages we need to compute? Write them down in terms of the variables $\mathbf{P}, \mathbf{Q}, \mathbf{R}$.

6. Use these messages to implement the max-product algorithm to infer for given $\mathbf{v}, \mathbf{P}, \mathbf{Q}, \mathbf{R}$:
$$\mathbf{h}^* = \underset{\mathbf{h} \in \{1,\ldots,H\}^T}{\arg\max}\; p(\mathbf{v}, \mathbf{h}). \tag{7}$$
(If that poses troubles, you can also implement the brute-force variant, with T for-loops over the states $1, \ldots, H$. It may also help you to debug your max-product implementation.)

7. What is the most likely $\mathbf{h}$ for each of the two sequences $6, 2, 2, 4, 4, 6, 3$ and $5, 3, 2, 7, 6, 4, 6, 4$?