

Machine Learning A

708.064 13W 1sst KU

Exercises

Problems marked with * are optional.

1 Conditional Independence

- a) [1 P] For the probability distribution $P(A, B, C, D)$ with the factorization

$$P(A, B, C, D) = P(A)P(B)P(C|A, B)P(D|C)$$

show that the following conditional independence assumptions hold.

(i) $A \perp B | \emptyset$

(ii) $A \perp D | C$

- b) [1 P] For the probability distribution $P(A, B, C, D)$ with the factorization

$$P(A, B, C, D) = P(A)P(B|A)P(C|A)P(D|C)$$

show that the following conditional independence assumptions hold.

(i) $A \perp D | C$

(ii) $B \perp D | C$

- c) [1* P] Find a probability distribution $P(A, B, C)$ in form of a probability table which fulfills $A \perp B | \emptyset$ but not $A \perp B | C$. Proof that your probability distribution really fulfills the two criteria.

2 Bayesian Networks

- a) [1 P] Construct a Bayesian network which represents the conditional independence assumptions of the probability distributions from example 1 and 2.

- b) [2 P] Construct two different Bayesian network which encodes exactly the following conditional independence assumptions

$$A \perp C | B$$

$$A \perp D | B$$

$$C \perp D | B$$

- c) [2 P] A doctor gives a patient a drug dependent on their age and gender. The patient has a probability to recover depending on whether s/he receives the drug, how old s/he is and which gender the patient has. Additionally it is known that age and gender are conditional independent if nothing else is known from the patient.¹

- (i) Draw the Bayesian network which describes this situation.
- (ii) How does the factorized probability distribution look like?
- (iii) Write down the formula to compute the probability that a patient recovers given that you know if s/he gets the drug. Write down the formula using only probabilities which are part of the factorized probability distribution.

3 d-separation

- a) [2 P] In Figure 1 you can see a Bayesian network which is used for the diagnosis of lung cancer and tuberculosis. Check if the following conditional independence assumptions are true or false.²

- (i) *tuberculosis* \perp *smoking* | *dyspnea*
- (ii) *lungcancer* \perp *bronchitis* | *smoking*
- (iii) *visit to Asia* \perp *smoking* | *lungcancer*
- (iv) *visit to Asia* \perp *smoking* | *lungcancer, dyspnea*

¹ Adapted from exercise 3.9 in D. Barber: Bayesian Reasoning and Machine Learning, Cambridge University Press, 2012.

²Exercise 3.3 in D. Barber: Bayesian Reasoning and Machine Learning, Cambridge University Press, 2012.

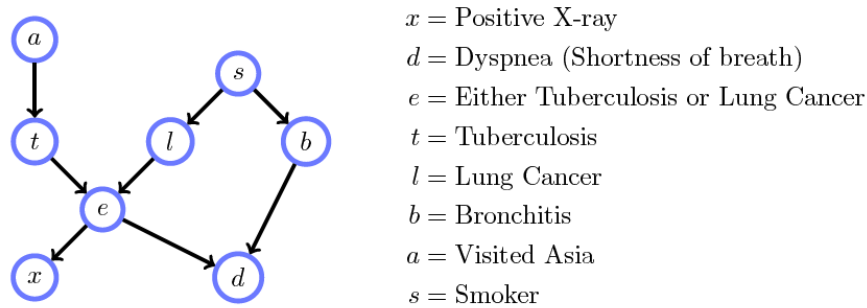


Figure 1: Bayesian network for the diagnosis of lung cancer and tuberculosis.

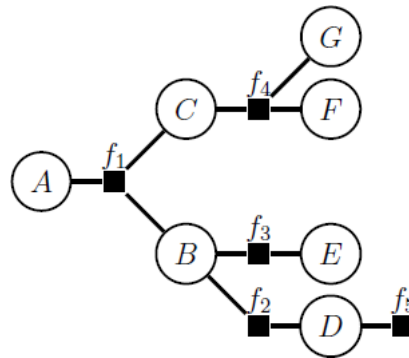


Figure 2: Factor graph for example 4

4 Inference in Factor Graphs

[2 P] Show how the given marginal probabilities can be computed in the factor graph in Fig. 2 using the sum-product algorithm. Start with drawing arrows to the factor graph which show how the messages are passed. After that write down how each of the messages can be computed and finally write down the formula for the marginal distribution.

a) $P(A)$

b) $P(B)$

5 Factor graphs: HMM

[5 P] Implement the sum-product algorithm for factor graphs in MATLAB to infer the hidden states of a HMM for the following problem.

An agent is located randomly in a 10×10 grid world. In each of $T = 20$ time steps he either stays at his current location with a probability of 25% or

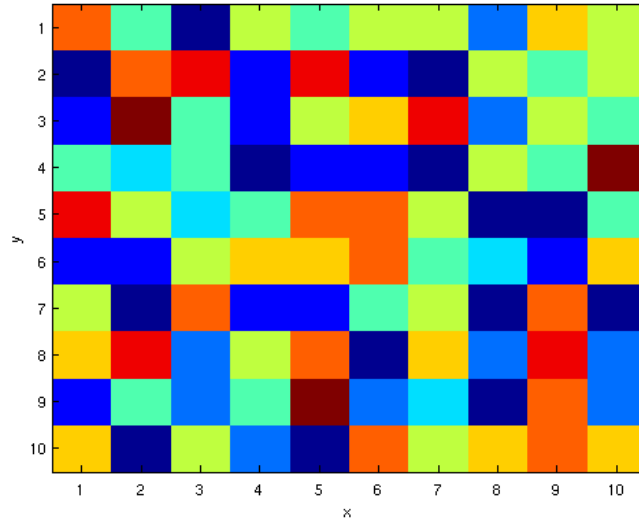


Figure 3: Grid world.

moves up, down, left or right, where each of these four actions is chosen with a probability of 18.75%. Each cell in the grid world is painted with a certain color that is chosen randomly from a set of k possible colors as shown in Fig. 3. This color is observed by the agent at each time step and reported to us. Only these color values are available to infer the initial location and the subsequent positions of the agent resulting in the Hidden Markov model (HMM) illustrated in Fig. 4.

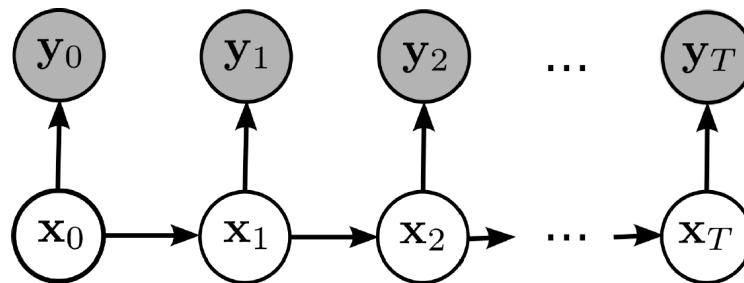


Figure 4: Hidden Markov model (HMM).

Modify the file `hmmMessagePassingTask.m` available for download on the course homepage ³ and implement the sum-product algorithm to solve this inference problem. Investigate and discuss the dependence of the solution on the number of different colors in the grid world (k). Hand in figures of representative results that show the actual agent trajectories, the most likely trajectories

³http://www.igi.tugraz.at/lehre/intern/MLA_WS1314_HW5.zip

and the probabilities of the agent positions at each time step as inferred by the sum-product algorithm.

Present your results clearly, structured and legible. Document them in such a way that anybody can reproduce them effortlessly. Send the code of your solution to florian.hubner@igi.tugraz.at

6 Markov Networks

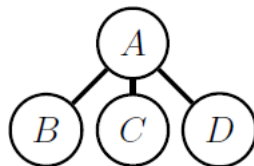
- a) [2 P] For the given joint probability draw an appropriate Markov network and write down the formula for the given conditional probability. Simplify the formula for the conditional probability as far as possible.

(i) $P(A, B, C, D, E) = \frac{1}{Z} \phi(A, C) \phi(B, C) \phi(C, D) \phi(C, E),$
 $P(A|C) = ?$

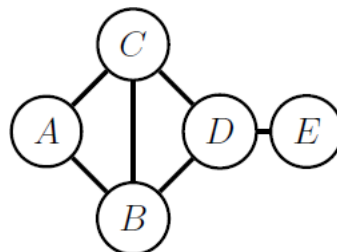
(ii) $P(A, B, C, D, E) = \frac{1}{Z} \phi(A, B, C) \phi(C, D) \phi(D, E),$
 $P(A|D) = ?$

- b) [3 P] In this example you have to write down the joint distribution for the given Markov network and proof that a given independence assumption holds.

(i) $B \perp C | A$



(ii) $A \perp E | D$



7 Junction Trees

[3 P] Consider the following distribution ⁴

$$P(A, B, C, D, E, F, G, H, I) = P(A)P(B|A)P(C|A)P(D|A)P(E|B)P(F|C)P(G|D)P(H|E, F)P(I|F, G)$$

- Draw the Bayesian network for this distribution.
- Draw the moralised graph.
- Draw the triangulated graph. Your triangulated should contain cliques of the smallest size possible.
- Draw a junction tree for the above graph and verify that it satisfies the running intersection property.
- Write down a suitable initialization of clique potentials.
- Find an appropriate message updating schedule.

8 Parameter Learning: Beta distribution

[2 P] You want to estimate the bias of a coin using a Bayesian approach. For that you toss the coin ten times and get the following result

$$\{h, t, t, t, h, t, h, t, t, t\}$$

- Compute the parameters of the posterior probability using a uniform Beta distribution as prior. Plot the posterior probability density function using the MATLAB function `betapdf`.
- You do another experiment with ten tosses and this time you get the following result

$$\{h, h, t, h, t, t, h, t, t, h\}$$

Again compute the parameters of the posterior probability with (i) a uniform Beta distribution (ii) using the parameters you obtained after the first experiment. Plot the two posterior probability density function.

- Compute the probability that the next toss will be h using the results from b).
- Explain the different results you get in b) and c).

⁴Exercise 6.4 in D. Barber: Bayesian Reasoning and Machine Learning, Cambridge University Press, 2012.

9 Beta distribution

[2* P] Show that the mean, variance, and mode of the beta distribution are given respectively by

$$\begin{aligned}E[\mu] &= \frac{a}{a+b} \\var[\mu] &= \frac{ab}{(a+b)^2(a+b+1)} \\mode[\mu] &= \frac{a-1}{a+b-2}.\end{aligned}$$

10 Parameter Learning: Naive Bayes Classifiers

[5 P] Implement an algorithm for learning a naive Bayes classifier and apply it to a spam email data set. You are required to use MATLAB for this assignment. The spam dataset is available for download on the course homepage⁵.

- a) Write a function called `nbayes_learn.m` that takes a training dataset for a binary classification task with binary attributes and returns the posterior Beta distributions of all model parameters (specified by variables a'_i and b'_i for the i th model parameter) of a naive Bayes classifier given a prior Beta distribution for each of the model parameters (specified by variables a_i and b_i for the i th model parameter).
- b) Write a function called `nbayes_predict.m` that takes a set of test data vectors and returns the most likely class label predictions for each input vector based on the posterior parameter distributions obtained in a).
- c) Use both functions to conduct the following experiment. For your assignment you will be working with a data set that was created a few years ago at the Hewlett Packard Research Labs as a testbed data set to test different spam email classification algorithms.
 - (i) Train a naive Bayes model on the first 2500 samples (using Laplace uniform prior distributions) and report the classification error of the trained model on a test data set consisting of the remaining examples that were not used for training.
 - (ii) Repeat the previous step, now training on the first $\{10, 50, 100, 200, \dots, 500\}$ samples, and again testing on the same test data as used in point 1 (samples 2501 through 4601). Report the classification error on the test dataset as a function of the number of training examples. Hand in a plot of this function.

⁵http://www.igi.tugraz.at/lehre/intern/MLA_WS1314_HW10.zip

- (iii) Comment on how accurate the classifier would be if it would randomly guess a class label or it would always pick the most common label in the training data. Compare these performance values to the results obtained for the naive Bayes model.

Present your results clearly, structured and legible. Document them in such a way that anybody can reproduce them effortlessly. Send the code of your solution to florian.hubner@igi.tugraz.at

11 K-means: Image compression

[3 P] Apply the k-means algorithm for lossy image compression by means of vector quantization.

- Download the 512×512 image `mandrill.tif`.⁶ Each pixel represents a point in a three dimensional (r,g,b) color space. Each color dimension encodes the corresponding intensity with an 8 bit integer.
- Cluster the pixels in color space using k-means with $k \in \{2, 4, 8, 16, 32, 64, 128\}$ clusters and replace the original color values with the indices of the closest cluster centers. You can use the MATLAB function `kmeans` to do the clustering. Determine the compression factor for each value of k and relate it to the quality of the image. Apply an appropriate quality measure of your choice.

Present your results clearly, structured and legible. Document them in such a way that anybody can reproduce them effortlessly. Send the code of your solution to florian.hubner@igi.tugraz.at

⁶http://www.igi.tugraz.at/lehre/intern/MLA_WS1314_HW11.zip

12 EM Algorithm for Gaussian Mixture Model

[5 P] In this task you have to implement the EM algorithm for Mixture of Gaussians. The algorithm can be found in the slides or in Bishop p.438,439.

1. Write the function `em` which takes the dataset and the number of clusters K as arguments. Make sure that you initialize the parameters $(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \pi_k)$ appropriately.
2. Download the dataset provided on the homepage⁷. It contains the complete dataset where \mathbf{x} contains the data and \mathbf{z} contains the correct classes. You can use \mathbf{z} to check if the correct clusters are found.

Let the algorithm run several times with $K = 5$ but with different initial cluster centers to see how it performs. Show some plots and describe the behaviour of the algorithm. Does it always find the correct clusters? Does the algorithm have any problems?

3. Now use different values for K and analyze what the algorithm does for K being lower or bigger than the correct amount of clusters.

Present your results clearly, structured and legible. Document them in such a way that anybody can reproduce them effortlessly. Send the code of your solution to florian.hubner@igi.tugraz.at

13 EM Algorithm for Mixture of Lines

Assume that the training examples $\mathbf{x}_n \in \mathbb{R}^2$ with $n = 1, \dots, N$ were generated from a mixture of K lines

$$P(x_{n,2} | z_{n,k} = 1) = \mathcal{N}(x_{n,2} | \theta_{k,1}x_{n,1} + \theta_{k,2}, \sigma_k) \quad (1)$$

where

$$\mathcal{N}(x | \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \quad (2)$$

and the hidden variable $z_{n,k} = 1$ if \mathbf{x}_n is generated from line k and 0 otherwise.

1. [1* P] Derive the update equations for the M-step of the EM algorithm for the variables θ_k and σ_k .
2. [2* P] Implement the EM algorithm for Mixture of Lines using the update equations you derived in 1. Use the provided dataset⁸ to evaluate your implementation. Show some plots of intermediate steps and describe what is happening.

⁷http://www.igi.tugraz.at/lehre/intern/MLA_WS1314_HW12.zip

⁸http://www.igi.tugraz.at/lehre/intern/MLA_WS1314_HW13.zip

Present your results clearly, structured and legible. Document them in such a way that anybody can reproduce them effortlessly. Send the code of your solution to florian.hubner@igi.tugraz.at