Probabilistic Models: Spring 2013
Exercise session 5 (exercises 19–23)

Instructions: All course participants are requested to submit their exercise solutions as follows:

- Deadline: before 12 o'clock noon of the Wednesday when the corresponding exercise session will be held
- Submission as a PDF file by email to joonas.paalasmaa at helsinki.fi (cc: to petri.myllymaki at cs.helsinki.fi)
- If you write the solutions by hand, please scan your paper. However, we strongly recommend that you type your solution by using a word processor. LaTeX is of course especially suitable for typesetting math, and it also has a convenient front-end LyX.
- Use as the title of your paper and subject of the email:
  "ProMo-2013, Exercise session n, yourlastname"
- Use as the file name:
  "ProMo-2013, Exercise session n, yourlastname.pdf"
- In all the exercises, do not just give the answer, but also the derivation of how you obtained it.
- Participants are encouraged to write computer programs to derive solutions whenever appropriate. In this case, please enclose the program source code too as a separate file.

After the exercise session, you are allowed to send a modified version of one of the solutions you sent before the exercise session:

- You can only modify a solution that you submitted before the 12 o'clock deadline.
- You can only send a modified version for one solution, and only if you attended the exercise session.
- Deadline: midnight after the exercise session.
- Submission as before, but this time please send only the modified solution, not the other (unchanged) solutions. Enclose the original solution first and then continue with the new material: first explain what you did wrong in the first time and then continue with modifications.
- As the title of the submission, please use:
  "ProMo-2013, Exercise session n, yourlastname, modified exercise x."

19. Consider the following training data set D, which was constructed by taking the 10 most probable vectors from the list produced in part 14(c) of the last week's exercise, and assigning them to the most probable class:

X_1	X_2	X_3	X_4	Y
0	0	0	1	1
0	1	0	0	2
0	1	0	1	1
0	1	1	0	2
1	0	0	0	1
1	0	0	1	1
1	0	1	1	1
1	1	0	0	3
1	1	0	1	1
1	1	1	0	2

Now based on this data, learn a Naive Bayes classifier by:

a) Using the maximum likelihood parameters.

b) Using the expected parameters with the BDeu prior using equivalence sample size of 1.

20. Using the two Naive Bayes classifiers (ML parameters/expected BDeu parameters) learned in the previous exercise:

a) Compute the classification distribution $P(Y \mid X_1, X_2, X_3, X_4)$ for each of the 16 cases (X_1, X_2, X_3, X_4).

b) Consider the following "test set" consisting of the 6 cases not found in the training set D, assuming that the correct class in each case is the one maximizing the classification probability given by the original generating NB classifier (Last week's exercise 14(a)):

X_1	X_2	X_3	X_4	Y
0	0	0	0	1
0	0	1	0	2
0	0	1	1	1
0	1	1	1	2
1	0	1	0	3
1	1	1	1	2

Compute the 0/1 loss in the "test set".

c) Compute the logarithmic loss in the same "test set" as in b).

Loss functions for one instance are as follows.

0/1 loss:

$$L_{0/1}(\hat{y}, y) = \begin{cases} 0 & \text{if } \hat{y} = y \\ 1 & \text{otherwise} \end{cases}$$

where $\hat{y}$ is the correct class and y is the predicted class.

Logarithmic loss:

$$L_{\log}(\hat{y}, p_{\hat{y}}) = -\log(p_{\hat{y}}),$$

where $p_{\hat{y}}$ is the prediction probability of class $\hat{y}$. Here we assume that $\log 0 = -\infty$.

The loss in the test set is the sum of individual losses.

21. Still using the same training data set D of size 10:

a) What is the marginal likelihood with the Naive Bayes structure (with BDeu priors, equivalent sample size 1)? Calculate it in 3 different ways to make sure they all produce the same result: using the "gamma-formula" directly, and using sequentially the predictive distribution with two different orderings of the data (e.g. take first any ordering and then reverse it).

b) What is the marginal likelihood with the empty graph (with BDeu priors, equivalent sample size 1)?

c) What should be the ratio of the prior probabilities of these two structures to make their posterior probabilities $P(M \mid D)$ equal?

22. Consider two binary variables X and Y. Prove that the structures $X \to Y$ and $Y \to X$ have always the same BDeu score (regardless of data and hyperparameters).

23. Consider three binary variables X, Y and Z. We have observed 100 data vectors. Counts for different vectors are as follows:

X	Y	Z	count
0	0	0	10
0	0	1	13
0	1	0	1
0	1	1	44
1	0	0	2
1	0	1	18
1	1	0	4
1	1	1	8

Given the data

a) Find a DAG that maximizes the marginal likelihood given BDeu priors with equivalent sample size 1. (Hint: BDeu score is likelihood equivalent)

b) Given the structure found in a), give the posterior distributions of the parameters given BDeu priors with equivalent sample size 1.

c) Given the structure found in a), find the expected parameters given BDeu priors with equivalent sample size 1.