HOMEWORK 1

Handed out: Tuesday, Jan. 23, 2018 Due: Monday, Feb. 5 midnight

Notes:

1. We encourage typed (Latex or Word) homework.
2. Submit a zipped directory with all files (not just the PDF) to pgm.notredame@gmail.com.Include your computer programs & data files and a Readme file providing scripts on how to run them.
3. Form of your Email attachment/subject: HW1_LastName_FirstName.zip (rar)

Problem 1. **Conditional Independence:** Consider three binary variables x, y and z with joint distribution as shown in Table 1. (a) Evaluate the distributions $p(x)$, $p(z|x)$ and $p(y|z)$, (b) Show that x and y are marginally dependent, but they become independent when conditioned on z. (c) Numerically, show that $p(x,y,z) = p(x)\,p(y|z)\,p(z|x)$ and draw the corresponding directed graph.

Table 1: Joint distribution for Example 1.

x	0	0	0	0	1	1	1	1
y	0	0	1	1	0	0	1	1
z	0	1	0	1	0	1	0	1
$p(x,y,z)$	0.192	0.144	0.048	0.216	0.192	0.064	0.048	0.096

Problem 2. **Plate representation:** We consider the relevance vector machine (a model for Bayesian regression with prior for the regression parameters that leads to sparse solutions). The model defines a conditional distribution for a real-valued target variable *t*, given an input vector **x**, which takes the form

$$p(t|\boldsymbol{x},\boldsymbol{w},\beta) = \mathcal{N}(t|y(\boldsymbol{x}),\beta^{-1})$$

where $\beta = \sigma^{-2}$ is the noise precision and the mean is given by a linear model of the form:

$$y = w^T\phi(\boldsymbol{x})$$

with fixed nonlinear basis functions $\phi_i(\boldsymbol{x})$, which will typically include a constant term so that the corresponding weight parameter represents a 'bias'.

Now suppose, we are given a set of N observations of the input vector $\mathbf{x}$, which we denote collectively by a data matrix $\mathbf{X}$ whose n^{th} row is $\boldsymbol{x}_n^T$ with $n = 1,\ldots,N$. The corresponding target values are given by $\mathbf{t} = (t_1,\ldots,t_N)^T$ and hence the likelihood function is given by:

$$p(\mathbf{t}|\mathbf{X},\mathbf{w},\beta) = \prod_{n=1}^{N} p(t_n|x_n,\mathbf{w},\beta) \tag{1}$$

We introduce a zero mean Gaussian prior distribution over the parameter vector $\mathbf{w}$:

$$p(\mathbf{w}|\boldsymbol{\alpha}) = \prod_{i=1}^{M} \mathcal{N}(w_i|0,\alpha_i) \tag{2}$$

where α_i denotes the hyperparameter associated with the weight parameter w_i. One can show that, when we maximize the evidence with respect to these hyperparameters, a significant proportion of them go to infinity, and the corresponding weight parameters have posterior distributions that are concentrated at zero. For details you can consult the reference given below.

Represent this two-equation Bayesian regression model (Eqs. (1) and (2)) by using plate representation of an appropriate PGM.

Reference: C. Bishop, Pattern Recognition and Machine Learning, Section 7.2.

Problem 3. Gaussian Graphical Model: Consider a directed graphical model as shown in Fig. 1. Assume that the network in Fig. 1 is a linear Gaussian graphical model and hence, conditionals given the parents are Gaussians with means linear on the parents. As shown in Lecture 4, the expectation and covariance follow the following recursive relations

$$\mathbb{E}[x_i] = \sum_{j \in pa_i} w_{ij}\mathbb{E}[x_j] + b_i$$

$$\operatorname{cov}[x_i, x_j] = \sum_{k \in pa_j} w_{jk}\operatorname{cov}[x_i, x_k] + I_{ij}v_j$$

where I_{ij} is the $(i,j)-$th element of an identity matrix. Use these recursive relations to compute the mean and covariance of the joint distribution.

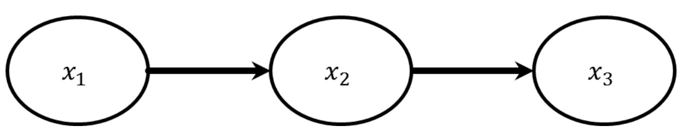


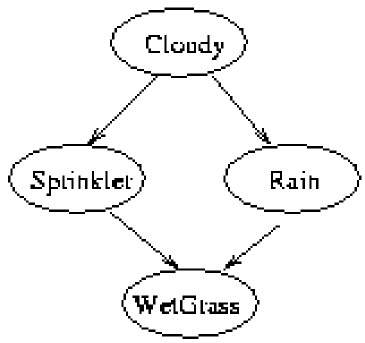
Fig. 1 Graphical model for problem 3.

Reference: C. Bishop, Pattern Recognition and Machine Learning, Section 8.1.4.

Problem 4. Bayesian network constructor: Write a code to construct a Bayesian Network (directed graphical model). The code should store (a) node/edge information, (b) conditional probability table corresponding to each node and (c) take user provided evidence.

Problem 5. Variable elimination: Write a code for variable elimination that should interact with the code developed in Problem 4. Use it for computing $p(\mathrm{S}|\mathrm{W{=}T})$ (see Fig. 2).

P(C=F)	P(C=T)
0.5	0.5



C	P(S=F)	P(S=T)
F	0.5	0.5
T	0.9	0.1

C	P(R=F)	P(R=T)
F	0.8	0.2
T	0.2	0.8

S	R	P(W=F)	P(W=T)
F	F	1.0	0.0
T	F	0.1	0.9
F	T	0.1	0.9
T	T	0.01	0.99

Fig. 2 Bayesian network for Problem 5