Coursework 1
Mathematics for Machine Learning (CO-496)

This coursework has only a writing component. The writing assignment requires plots, which you can create using any method of your choice. However, make sure that the figures are vector graphics, axes are legible, scales are appropriate etc. You should not submit the code used to create these plots.

No aspect of your submission may be hand-drawn. You are strongly encouraged to use LaTeXto create the written component. We will provide a template.

You are required to submit the following:

- A file `write_up.pdf` for your written answers.

1 Statistics and Probabilities

1. **[8 marks]** Compute the sample mean and the sample covariance matrix of the following dataset (use $1/N$ for the covariance matrix). Describe the computations you used to get to the answer.

$$\mathcal{D} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -4 \\ 4 \\ 2 \end{bmatrix} \right\}$$

2. **[9 marks]** Generate two datasets $\{(x_1, x_2)_{n=1}^N\}$ of $N = 100$ data points each. The datasets have mean

$$\boldsymbol{\mu} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

   and marginal variances $\sigma_1^2 = 2$, $\sigma_2^2 = 0.5$. Ensure that the shape of the datasets you generate is different. Visualize the two datasets and explain how you generated them so that their shapes are different.

3. **[27 marks]** Nora and Noah spent the summer on writing a computer program that solves AI. However, they encounter the problem that their code seems to be failing randomly when compiling. Nora and Noah want to estimate the probability of successful compilation using a probabilistic model. They assume that when compiling the code N times (without any changes to the code) gives i.i.d. results. Furthermore, the probability of success can be described by a Bernoulli distribution with an unknown parameter μ. As good Bayesians, they place a conjugate Beta prior on this unknown parameter, where the parameters of this beta prior are $\alpha = 2, \beta = 2$. They have now run $N = 20$ experiments, and 6 of them successfully compiles, and 14 failed.

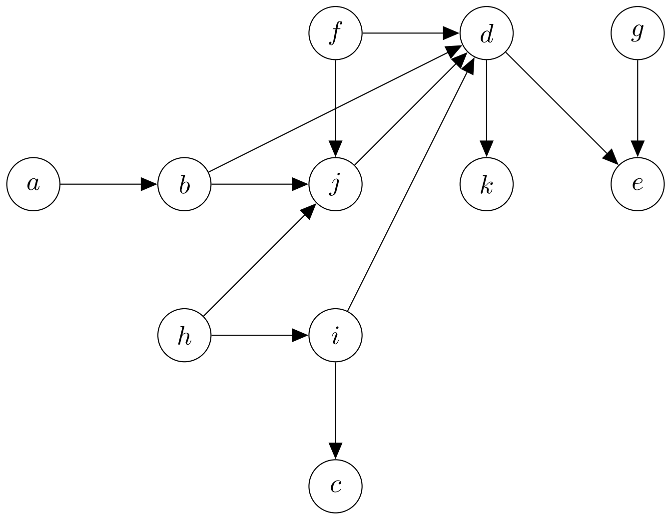
   - Compute the posterior distribution on μ (derive your result) and plot it.
   - What has changed from the prior to the posterior? Describe properties of the prior and the posterior.

2 Graphical Models

1. **[16 marks]** Given a factorized joint distribution, draw the corresponding directed graphical model (you can scan in a picture or use the tikz-bayesnet)

$$p(a, b, c, d, e, f) = p(a|b, c)p(c|b)p(d)p(e|d, a)p(f|c, d, e)p(b)$$

2. **[40 marks]** You are given the following graphical model:



Determine whether random variables are conditionally independent. Justify your answer.

Every correct answer gives +2, every incorrect answer -2 marks. The total mark for this question is lower-bounded by 0. Up to 40 marks.

(a) $a \perp\!\!\!\perp g | k$

(b) $d \perp\!\!\!\perp h | i$

(c) $g \perp\!\!\!\perp a | c$

(d) $e \perp\!\!\!\perp j | k$

(e) $b \perp\!\!\!\perp e | j$

(f) $j \perp\!\!\!\perp c | \{k, g\}$

(g) $a \perp\!\!\!\perp k$

(h) $a \perp\!\!\!\perp k | e$

(i) $h \perp\!\!\!\perp d | \{j, e\}$

(j) $b \perp\!\!\!\perp h | e$

(k) $h \perp\!\!\!\perp c | d$

(l) $a \perp\!\!\!\perp f | k$

(m) $i \perp\!\!\!\perp a | j$

(n) $g \perp\!\!\!\perp h | e$

(o) $g \perp\!\!\!\perp f | d$

(p) $g \perp\!\!\!\perp h$

(q) $a \perp\!\!\!\perp i | g$

(r) $b \perp\!\!\!\perp h$

(s) $b \perp\!\!\!\perp h | g$

(t) $i \perp\!\!\!\perp a$