

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 \LaTeX to 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

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

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 compile, and 14 failed.

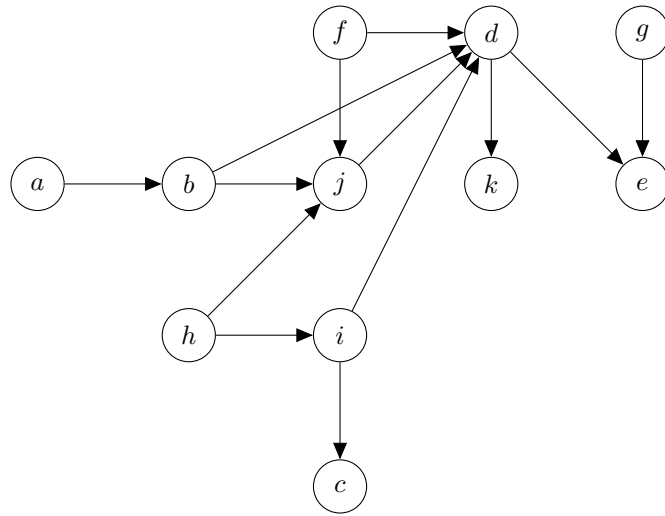
- Compute the posterior distribution on μ (derive your result) and plot it.
- What has changed from the prior to the posterior? Describe properties of the prior and the posterior.

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) $a \perp\!\!\!\perp g k$ | CI: e blocks the path as arrows meet head-to-head and neither node nor descendants, are conditioned upon |
| (b) $d \perp\!\!\!\perp h i$ | CD: $h \rightarrow j \rightarrow d$ is unblocked (arrows head to tail at j and j not observed) |
| (c) $g \perp\!\!\!\perp a c$ | CI: e blocks the path as arrows meet head-to-head and neither node nor descendants, are conditioned upon |
| (d) $e \perp\!\!\!\perp j k$ | CD: $j \rightarrow d \rightarrow e$ (arrows meet head to tail at d but d not observed) |
| (e) $b \perp\!\!\!\perp e j$ | CD: $b \rightarrow d \rightarrow e$ (j observed but alternative path directly from $b \rightarrow d$. d doesn't block the path as arrows meet head to tail at d but it isn't observed) |
| (f) $j \perp\!\!\!\perp c \{k, g\}$ | CD: $j \rightarrow h \rightarrow i \rightarrow c$ (arrows head-to-head at h but k is descendant of h so h unblocked; arrows tail to tail at i and i not observed) |
| (g) $a \perp\!\!\!\perp k$ | CD: $a \rightarrow b \rightarrow d \rightarrow k$ (all paths head to tail and no set observed so no paths blocked) |
| (h) $a \perp\!\!\!\perp k e$ | CD: $a \rightarrow b \rightarrow d \rightarrow k$ (all paths head to tail but none are blocked as e isn't on the path) |
| (i) $h \perp\!\!\!\perp d \{j, e\}$ | CD: $h \rightarrow i \rightarrow d$ (arrows head to tail at i but i not observed) |
| (j) $b \perp\!\!\!\perp h e$ | CD: j blocks the path as arrows arrive head-to-head at j and e is a descendant of j for j unblocked |
| (k) $h \perp\!\!\!\perp c d$ | CD: $h \rightarrow i \rightarrow c$ (arrows head to tail at i but i not observed) |
| (l) $a \perp\!\!\!\perp f k$ | CD: $a \rightarrow b \rightarrow j \rightarrow f$ (arrows head to tail at b and b not observed; arrows head to head at j but k is descendant of j so j unblocked) |
| (m) $i \perp\!\!\!\perp a j$ | CD: $i \rightarrow h \rightarrow j \rightarrow b \rightarrow a$ (h unblocked as head to head at h but j descendant of h; j unblocked as head to head but j observed; head to tail at b and b not observed) |
| (n) $g \perp\!\!\!\perp h e$ | CD: $h \rightarrow j \rightarrow d \rightarrow e \rightarrow g$ (head to tail at j and d and neither observed; head to head at e and e observed) |
| (o) $g \perp\!\!\!\perp f d$ | CI: e blocks path (head to head at e and e not observed and no descendants) |
| (p) $g \perp\!\!\!\perp h$ | CI: e blocks path (head to head at e and e not observed and no descendants) |
| (q) $a \perp\!\!\!\perp i g$ | CI: j & d block paths (j blocks $a \rightarrow h$ as head to head at j but j and descendants unobserved; d blocks $a \rightarrow i$ as head to head at d and d descendant of j (so unobserved also)) |
| (r) $b \perp\!\!\!\perp h$ | CI: j & d block paths (j blocks $b \rightarrow h$ directly as head to head at j and nothing observed; d blocks $b \rightarrow d \rightarrow i \rightarrow h$ as head to head at d and nothing observed) |

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

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

CI: j&d block paths (same reason as questions r) as g is not descendant of j or d)
CI: j & d block paths (j blocks $a \longrightarrow b \longrightarrow j \longrightarrow h \longrightarrow i$ as head to head at j and nothing observed; d blocks remaining paths as head to head at d and nothing observed)