Introduction to Probabilistic Graphical Models: Exercises

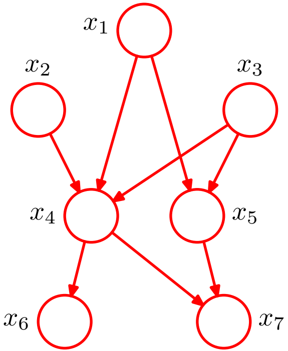
Cédric Archambeau

Xerox Research Centre Europe
cedric.archambeau@xrce.xerox.com

Pascal Bootcamp
Marseille, France, July 2010

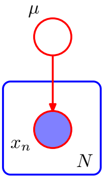
Exercise 1: basics

Given the following graphical model:

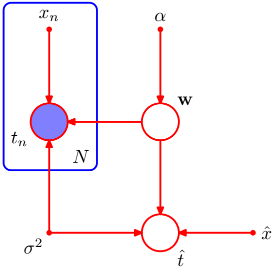


How can $p(x_1, \ldots, x_7)$ be docomposed?

Exercise 2: d-separation



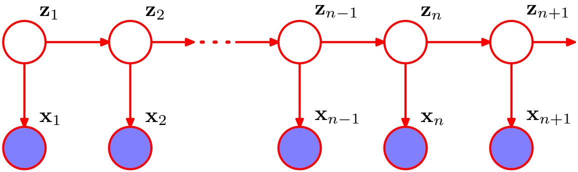
(a)



(b)

- Let $p(x|\mu)$ be a univariate Gaussian and $p(\mu)$ the prior on the mean. Assume we observe $\{x_n\}$. The corresponding Bayesian network is shown in Fig. a. Unfold the graphical model and write down the joint. Can we exploit any CI? What happens if we marginalise wrt μ?
- Fig. b is the graphical model for Bayesian polynomial regression. The variables $\{\mathbf{x}_n, t_n\}$ are the input-output pairs and $\mathbf{w}$ is the vector or regression weights. The variable $\hat{t}$ is the prediction. Are training outputs and prediction independent? What is the consequence in practice?

Exercise 3: linear dynamical system



- Consider a state space model with Gaussian transition density and Gaussian noise:

  $$p(\mathbf{z}_n|\mathbf{z}_{n-1}) = \mathcal{N}(\mathbf{F}\mathbf{z}_{n-1}, \mathbf{R}), \qquad p(\mathbf{x}_n|\mathbf{z}_n) = \mathcal{N}(\mathbf{H}\mathbf{z}_n, \mathbf{Q}).$$

  where $\mathbf{F}$, $\mathbf{H}$, $\mathbf{R}$ and $\mathbf{Q}$ are parameters.
- Write the factor graph and the joint density. Can we decompose $p(\mathbf{x}_{n+1}|\mathbf{x}_1, \ldots, \mathbf{x}_n)$? Motivate.
- Use the sum-product algorithm to derive the forward recursion (Kalman filter). Write the recursion for the message from f_1 to z_2. (Hint: first rewrite the message from h to f_1?) What specific form does it take for arbitrary n?
- Use Gaussian identities for write down the explicit form. Code and test the algorithm.

Multivariate Gaussian distribution

Let $\mathbf{x}$ be a D-dimensional Gaussian random vector.

The density of $\mathbf{x}$ is defined as

$$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-D/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\},$$

where $\boldsymbol{\mu} \in \mathbb{R}^{D \times 1}$ is the mean and $\boldsymbol{\Sigma} \in \mathbb{R}^{D \times D}$ is the covariance matrix.

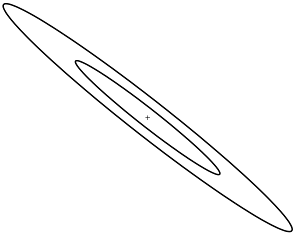


Figure: 2-dimensional Gaussian.

Gaussian identities

Consider the following two Gaussian distributions:

$$p(\mathbf{x}) = \mathcal{N}(\boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{xx}}),$$
$$p(\mathbf{y}|\mathbf{x}) = \mathcal{N}(\mathbf{A}\mathbf{x} + \mathbf{b}, \boldsymbol{\Lambda}).$$

The **marginal** $p(\mathbf{y})$ is Gaussian with mean and covariance given by

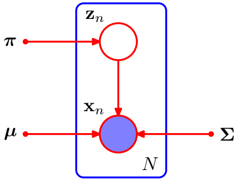
$$\boldsymbol{\mu}_{\mathbf{y}} = \mathbf{A}\boldsymbol{\mu}_{\mathbf{x}} + \mathbf{b},$$
$$\boldsymbol{\Sigma}_{\mathbf{yy}} = \boldsymbol{\Lambda} + \mathbf{A}\boldsymbol{\Sigma}_{\mathbf{xx}}\mathbf{A}^{\top}.$$

The **posterior** $p(\mathbf{x}|\mathbf{y})$ is Gaussian with mean and covariance equal to

$$\boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}} = \boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}}\{\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1}\boldsymbol{\mu}_{\mathbf{x}} + \mathbf{A}^{\top}\boldsymbol{\Lambda}^{-1}(\mathbf{y} - \mathbf{b})\},$$
$$\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}} = (\boldsymbol{\Sigma}_{\mathbf{xx}}^{-1} + \mathbf{A}^{\top}\boldsymbol{\Lambda}^{-1}\mathbf{A})^{-1}.$$

(For proofs see for example chapter 2 of Bishop, 2006.)

Exercise 4: Gaussian mixture model



- Define the Bayesian network and write down the log-likelihood. Show that the posterior $q(z)$ is of proportional to $\exp\{\ln p(\mathbf{x}, z; \boldsymbol{\theta})\}$.
- Write the expected log-complete likelihood. Derive the M step. Are there constraints to take into account?
- Download the wine data from the UCI repository and implement the algorithm. Investigate how the different components move and change shape (by plotting ellipses) as a function of the number of iterations. (Consider only two dimensions.)
- Split the data in two sets. Train the algorithm on one and compute the log-likelihood of both sets. Plot their evolution as a function of the number of components. What do you observe?

Solution 1 and 2

- 1: Bishop chapter 8, p 362 (top).
- 2: Bishop chapter 8, p 379 (bottom) and p 380 (top).

Solution 3

- Factor graph idem HMM but with Gaussian parametric form. No decomposition (check d-separation).
- Forward recursion idem HMM. First message is proportional to $p(z_1|x_1)$. Then:

$$\mu_{f_2 \to x_2}(z_2) = p(x_2|z_2) \int p(z_2, z_1|x_1) dz_1 = p(x_2|z_2) p(z_2|x_1) \propto p(z_2|x_2, x_1).$$

$$\mu_{f_n \to x_n}(z_n) \propto p(x_n|z_n) p(z_n|x_{n-1}, \dots, x_1) \propto p(z_n|x_n, \dots, x_1).$$

The $p(z_n|x_{n-1}, \dots, x_1)$ is called the predictive density. Direct application of first Gaussian identity. The second is found by completing the squares (second identity). It is called the filtering density. In practice you need to assume some Gaussian prior on the initial state.