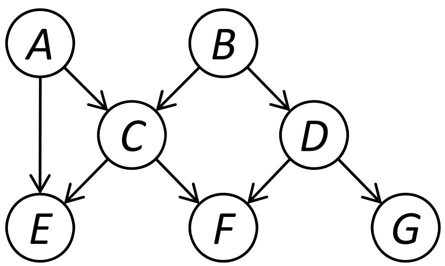
Homework set 3

(Lecture 5)

5.1. Bayesian networks in practice

From the following Bayesian network, compute *p*(*A*|*E,F*) in terms of known conditional probabilities:



$$p(A|E,F) = \frac{p(A,E,F)}{p(E,F)} = \frac{p(A,E,F)}{\sum_A p(A,E,F)}$$

$$p(A,E,F) = \sum_{B,C,D,G} p(A,B,C,D,E,F,G)$$

$$= \sum_{B,C,D,G} p(A)\,p(B)\,p(C|A,B)\,p(D|B)\,p(E|A,C)\,p(F|C,D)\,p(G|D)$$

$$= \sum_{B,C,D} p(A)\,p(B)\,p(C|A,B)\,p(D|B)\,p(E|A,C)\,p(F|C,D)$$

$$= p(A)\sum_B p(B)\sum_C p(C|A,B)\,p(E|A,C)\sum_D p(D|B)\,p(F|C,D)$$

5.2 Laplace approximation

a) Prove the Laplace approximation:

$$p(D \mid M) \approx p\left(D \mid M, \hat{\theta}_{\text{MAP}}\right) p\left(\hat{\theta}_{\text{MAP}} \mid M\right) \frac{1}{\sqrt{\det \frac{\mathbf{H}}{2\pi}}}$$

with

$$\mathbf{H} = -\nabla\nabla \log p\left(\theta \mid D, M\right)\Big|_{\theta=\hat{\theta}_{\text{MAP}}}$$

b) What is **H** when the posterior over θ is a multivariate Gaussian with mean $\hat{\theta}_{\text{MAP}}$ and covariance matrix **Σ**?

$$p(D \mid M) = \int p(D \mid M, \theta)\, p(\theta \mid M)\, d\theta$$

Unnormalized probability distribution over θ:

$$p(D \mid M, \theta)\, p(\theta \mid M)$$

Expand its log around $\hat{\theta}_{\mathrm{MAP}}$:

$$\log\left[p(D \mid M, \theta)\, p(\theta \mid M) \right] \approx \log\left[p\left(D \mid M, \hat{\theta}_{\mathrm{MAP}}\right) p\left(\hat{\theta}_{\mathrm{MAP}} \mid M\right) \right] - \frac{1}{2}\left(\theta - \hat{\theta}_{\mathrm{MAP}}\right)^{\mathrm{T}} \mathbf{H} \left(\theta - \hat{\theta}_{\mathrm{MAP}}\right)$$

$$\mathbf{H} = -\nabla\nabla \log\left[p(D \mid M, \theta)\, p(\theta \mid M) \right]\Big|_{\theta = \hat{\theta}_{\mathrm{MAP}}} = -\nabla\nabla \log p(\theta \mid D, M)\Big|_{\theta = \hat{\theta}_{\mathrm{MAP}}}$$

$$p(D \mid M, \theta)\, p(\theta \mid M) \approx p\left(D \mid M, \hat{\theta}_{\mathrm{MAP}}\right) p\left(\hat{\theta}_{\mathrm{MAP}} \mid M\right) \exp\left[-\frac{1}{2}\left(\theta - \hat{\theta}_{\mathrm{MAP}}\right)^{\mathrm{T}} \mathbf{H} \left(\theta - \hat{\theta}_{\mathrm{MAP}}\right) \right]$$

Normalization factor:

$$p(D \mid M) \approx p\left(D \mid M, \hat{\theta}_{\mathrm{MAP}}\right) p\left(\hat{\theta}_{\mathrm{MAP}} \mid M\right) \int \exp\left[-\frac{1}{2}\left(\theta - \hat{\theta}_{\mathrm{MAP}}\right)^{\mathrm{T}} \mathbf{H} \left(\theta - \hat{\theta}_{\mathrm{MAP}}\right) \right] d\theta$$

$$= p\left(D \mid M, \hat{\theta}_{\mathrm{MAP}}\right) p\left(\hat{\theta}_{\mathrm{MAP}} \mid M\right) \frac{1}{\sqrt{\det \frac{\mathbf{H}}{2\pi}}} = p\left(D \mid M, \hat{\theta}_{\mathrm{MAP}}\right) p\left(\hat{\theta}_{\mathrm{MAP}} \mid M\right) \sqrt{\frac{(2\pi)^d}{\det \mathbf{H}}}$$

Multivariate Gaussian

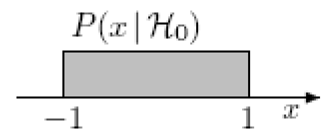
$$p\left(\theta \mid D, M\right) = \frac{1}{\sqrt{\left(2\pi\right)^{d} \det \mathbf{\Sigma}}} \exp\left[-\frac{1}{2}\left(\theta - \hat{\theta}_{\mathrm{MAP}}\right)^{\mathrm{T}} \mathbf{\Sigma}^{-1} \left(\theta - \hat{\theta}_{\mathrm{MAP}}\right)\right]$$

$$\mathbf{H} = -\nabla\nabla \log p\left(\theta \mid D, M\right)\Big|_{\theta=\hat{\theta}_{\mathrm{MAP}}} = \mathbf{\Sigma}^{-1}$$

5.3 Bayesian model comparison 1

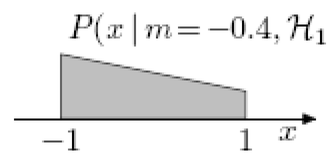
Exercise 28.1.[3] Random variables x come independently from a probability distribution $P(x)$. According to model $\mathcal{H}_0$, $P(x)$ is a uniform distribution

$$P(x \mid \mathcal{H}_0) = \frac{1}{2} \qquad x \in (-1, 1). \qquad (28.20)$$



According to model $\mathcal{H}_1$, $P(x)$ is a nonuniform distribution with an unknown parameter $m \in (-1, 1)$:

$$P(x \mid m, \mathcal{H}_1) = \frac{1}{2}(1 + mx) \qquad x \in (-1, 1). \qquad (28.21)$$



Given the data $D = \{0.3, 0.5, 0.7, 0.8, 0.9\}$, what is the evidence for $\mathcal{H}_0$ and $\mathcal{H}_1$?

Guesses?

$$p(D \mid H_0) = \prod_{i=1}^{5} p(x_i \mid H_0) = \prod_{i=1}^{5} \frac{1}{2} = \frac{1}{2^5}$$

$$
\begin{aligned}
p\left(D \mid H_1\right) &= \int p\left(D \mid m, H_1\right) p\left(m \mid H_1\right) dm \\
&= \int \left(\prod_{i=1}^{5} p\left(x_i \mid m, H_1\right) \right) p\left(m \mid H_1\right) dm \\
&= \int_{-1}^{1} \left(\prod_{i=1}^{5} \left(\frac{1}{2}\left(1 + m x_i\right) \right) \right) \frac{1}{2} dm \\
&= \frac{1}{2^6} \int_{-1}^{1} \prod_{i=1}^{5} \left(1 + m x_i\right) dm \\
&= \frac{1}{2^6} \int_{-1}^{1} \left(1 + m^2 \sum_{i>j} x_i x_j + m^4 \left(\prod_{i=1}^{5} x_i \right) \sum_{i=1}^{5} \frac{1}{x_i} \right) dm \\
&= \frac{1}{2^6} \left(2 + \frac{2}{3} \sum_{i>j} x_i x_j + \frac{2}{5} \left(\prod_{i=1}^{5} x_i \right) \sum_{i=1}^{5} \frac{1}{x_i} \right) \\
&= \frac{1}{2^5} \left(1 + \frac{1}{3} \sum_{i>j} x_i x_j + \frac{1}{5} \left(\prod_{i=1}^{5} x_i \right) \sum_{i=1}^{5} \frac{1}{x_i} \right)
\end{aligned}
$$

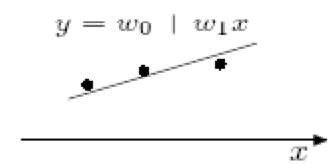
$$
\frac{p\left(D \mid H_1\right)}{p\left(D \mid H_0\right)} = 1 + \frac{1}{3} \sum_{i>j} x_i x_j + \frac{1}{5} \left(\prod_{i=1}^{5} x_i \right) \sum_{i=1}^{5} \frac{1}{x_i} \approx 2.46
$$

5.4 Bayesian model comparison 2

Exercise 28.2.[3] Datapoints (x, t) are believed to come from a straight line. The experimenter chooses x, and t is Gaussian-distributed about

$$y = w_0 + w_1 x \qquad (28.22)$$

with variance σ_ν^2. According to model $\mathcal{H}_1$, the straight line is horizontal, so $w_1 = 0$. According to model $\mathcal{H}_2$, w_1 is a parameter with prior distribution Normal$(0, 1)$. Both models assign a prior distribution Normal$(0, 1)$ to w_0. Given the data set $D = \{(-8, 8), (-2, 10), (6, 11)\}$, and assuming the noise level is $\sigma_\nu = 1$, what is the evidence for each model?



Guesses?

$$p(D \mid H_1) = \int p(D \mid w_0, H_1)\, p(w_0 \mid H_1)\, dw_0$$

$$= \int \left(\prod_{i=1}^{N} p(x_i, t_i \mid w_0, H_1) \right) p(w_0 \mid H_1)\, dw_0$$

$$= \left(\prod_{i=1}^{N} p(x_i) \right) \int \left(\prod_{i=1}^{N} p(t_i \mid x_i, w_0, H_1) \right) p(w_0 \mid H_1)\, dw_0$$

$$= \left(\prod_{i=1}^{N} p(x_i) \right) \int \left(\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(t_i - w_0)^2} \right) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} w_0^2}\, dw_0$$

$$p\left(D \mid H_2\right)=\iint p\left(D \mid w_0, w_1, H_2\right) p\left(w_0, w_1 \mid H_2\right) dw_0 dw_1$$

$$=\iint\left(\prod_{i=1}^{N} p\left(x_i, t_i \mid w_0, w_1, H_2\right)\right) p\left(w_0, w_1 \mid H_2\right) dw_0 dw_1$$

$$=\left(\prod_{i=1}^{N} p\left(x_i\right)\right)\iint\left(\prod_{i=1}^{N} p\left(t_i \mid x_i, w_0, w_1, H_2\right)\right) p\left(w_0 \mid H_2\right) p\left(w_1 \mid H_2\right) dw_0 dw_1$$

$$=\left(\prod_{i=1}^{N} p\left(x_i\right)\right)\iint\left(\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(t_i-w_0-w_1 x_i\right)^2}\right) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}w_0^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}w_1^2} dw_0 dw_1$$

$$\frac{p\left(D \mid H_2\right)}{p\left(D \mid H_1\right)}=\frac{1}{\sqrt{\left(N+1\right)c_{xx}}} e^{\frac{\left(N+1\right)c_{xy}^2}{2c_{xx}}}$$

$$c_{xx}=\frac{1}{N+1}\sum_{i=1}^{N} x_i^2-\left(\frac{1}{N+1}\sum_{i=1}^{N} x_i\right)^2$$

$$c_{xy}=\frac{1}{N+1}\sum_{i=1}^{N} x_i y_i-\frac{1}{\left(N+1\right)^2}\left(\sum_{i=1}^{N} x_i\right)\left(\sum_{i=1}^{N} y_i\right)$$

$$\frac{p\left(D \mid H_2\right)}{p\left(D \mid H_1\right)} \approx 0.183$$

Any questions about the small project?