Solutions to Quiz in Bayesian Statistical Learning, AMSI Summer School 2027

Aishwarya Bhaskaran, University of New South Wales
Matias Quiroz, University of Technology Sydney

Problem 1

The density for X is $p_X(x) = \exp\left(-x^2/2\right)/\sqrt{2\pi}$. Since $y = h(x)$ and $h(x) = \exp(x)(1+\exp(x))$ one-to-one, by the transformation of random variables theorem with $h^{-1}(y) = \log(y/(1-y))$ and $\frac{d}{dy}h^{-1}(y) = 1/y + 1/(1-y) = 1/(y(1-y))$,

$$\begin{aligned} p_Y(y) &= p_X\left(h^{-1}(y)\right)\left|\frac{dh^{-1}(y)}{dy}\right| \\ &= \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\left(\log\left(\frac{y}{(1-y)}\right)\right)^2\right)\frac{1}{y(1-y)}, \quad 0 < y < 1. \end{aligned}$$

Problem 2

(a.)

First expand

$$\begin{aligned} a(x-h)^2 + k &= a(x^2 - 2hx + h^2) + k \\ &= ax^2 - a2hx + ah^2 + k. \end{aligned}$$

Matching this to

$$x^2 + 10x - 3,$$

we find $a = 1, h = -5$ and $k = -28$.

(b.)

This is the matrix extension of the idea above. First expand

$$\begin{aligned} (\mathbf{x}-\mathbf{h})^\top\mathbf{A}(\mathbf{x}-\mathbf{h}) + k &= (\mathbf{x}^\top - \mathbf{h}^\top)\mathbf{A}(\mathbf{x}-\mathbf{h}) + k \\ &= (\mathbf{x}^\top\mathbf{A} - \mathbf{h}^\top\mathbf{A})(\mathbf{x}-\mathbf{h}) + k \\ &= \mathbf{x}^\top\mathbf{A}\mathbf{x} - \mathbf{x}^\top\mathbf{A}\mathbf{h} - \mathbf{h}^\top\mathbf{A}\mathbf{x} + \mathbf{h}^\top\mathbf{A}\mathbf{h} + k. \end{aligned} \tag{1}$$

Note that all terms are scalars and that $a = a^\top$ for any scalar a. Thus, together with the symmetry of $\mathbf{A}$ ($\mathbf{A} = \mathbf{A}^\top$) and rules for transpose of products give

$$\begin{aligned}
\mathbf{x}^\top \mathbf{A}\mathbf{h} &= (\mathbf{x}^\top \mathbf{A}\mathbf{h})^\top \\
&= \mathbf{h}^\top (\mathbf{x}^\top \mathbf{A})^\top \\
&= \mathbf{h}^\top \mathbf{A}^\top (\mathbf{x}^\top)^\top \\
&= \mathbf{h}^\top \mathbf{A}^\top (\mathbf{x}^\top)^\top \\
&= \mathbf{h}^\top \mathbf{A}\mathbf{x},
\end{aligned}$$

and (1) becomes

$$\mathbf{x}^\top \mathbf{A}\mathbf{x} - \mathbf{x}^\top 2\mathbf{A}\mathbf{h} + \mathbf{h}^\top \mathbf{A}\mathbf{h} + k.$$

Matching this to

$$\mathbf{x}^\top \mathbf{A}\mathbf{x} + \mathbf{x}^\top \mathbf{b} + c,$$

we find $\mathbf{b} = -2\mathbf{A}\mathbf{h}$ and $c = \mathbf{h}^\top \mathbf{A}\mathbf{h} + k$. Solving for $\mathbf{h}$ and k

$$\mathbf{h} = -\frac{1}{2}\mathbf{A}^{-1}\mathbf{b},$$

and

$$\begin{aligned}
k &= c - \mathbf{h}^\top \mathbf{A}\mathbf{h} \\
&= c - \left(\frac{1}{2}\mathbf{A}^{-1}\mathbf{b}\right)^\top \mathbf{A}\frac{1}{2}\mathbf{A}^{-1}\mathbf{b} \\
&= c - \frac{1}{4}\mathbf{b}^\top (\mathbf{A}^{-1})^\top \mathbf{A}\mathbf{A}^{-1}\mathbf{b} \\
&= c - \frac{1}{4}\mathbf{b}^\top \mathbf{A}^{-1}\mathbf{b}
\end{aligned}$$

since $\mathbf{A}^{-1}$ is also symmetric.

Problem 3

(a.)

Repeated application of $\Pr(E \cap H) = \Pr(E|H)\Pr(H)$, or its conditional version $\Pr(E \cap H|B) = \Pr(E|B, H)\Pr(H|B)$ gives

$$\begin{aligned}
\Pr(\underbrace{A_4 \cap A_3 \cap A_2}_{E} \cap \underbrace{A_1}_{H}) &= \Pr(\underbrace{A_4 \cap A_3}_{E} \cap \underbrace{A_2}_{H} \mid \underbrace{A_1}_{B})\Pr(A_1) \\
&= \Pr(A_4 \cap A_3 | A_1 \cap A_2)\Pr(A_2|A_1)\Pr(A_1) \\
&= \Pr(A_4 | A_1 \cap A_2 \cap A_3)\Pr(A_3|A_1, A_2)\Pr(A_2|A_1)\Pr(A_1).
\end{aligned}$$

(b.)

By definition, $\Pr(E|H) = \Pr(E)$ if E and H are independent. The results then follows from above.

Problem 4

For a normal density

$$\begin{aligned}\mathcal{N}(x|\mu_n,\sigma_n^2) &\propto \exp\left(-\frac{1}{2\sigma_n^2}(x-\mu_n)^2\right)\\ &\propto \exp\left(-\frac{1}{2\sigma_n^2}x^2+\frac{\mu_n}{\sigma_n^2}x\right)\end{aligned} \tag{2}$$

.

Since

$$\begin{aligned}\mathcal{N}(x|\mu_1,\sigma_1^2)\mathcal{N}(x|\mu_2,\sigma_2^2) &\propto \exp\left(-\frac{1}{2\sigma_1^2}(x-\mu_1)^2\right)\exp\left(-\frac{1}{2\sigma_2^2}(x-\mu_2)^2\right)\\ &\propto \exp\left(-\frac{1}{2\sigma_1^2}\left(x^2-2\mu_1 x\right)-\frac{1}{2\sigma_2^2}\left(x^2-2\mu_2 x\right)\right)\\ &\propto \exp\left(-\frac{\left(\sigma_2^2+\sigma_1^2\right)}{2\sigma_1^2\sigma_2^2}x^2+\frac{\left(\sigma_2^2\mu_1+\sigma_1^2\mu_2\right)}{\sigma_1^2\sigma_2^2}x\right),\end{aligned}$$

by matching (2) we get

$$\begin{aligned}\frac{1}{\sigma_n^2} &= \frac{\left(\sigma_2^2+\sigma_1^2\right)}{\sigma_1^2\sigma_2^2}=\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\\ \frac{\mu_n}{\sigma_n^2} &= \frac{\left(\sigma_2^2\mu_1+\sigma_1^2\mu_2\right)}{\sigma_1^2\sigma_2^2}=\frac{1}{\sigma_1^2}\mu_1+\frac{1}{\sigma_2^2}\mu_2.\end{aligned}$$

Thus

$$\mathcal{N}(x|\mu_1,\sigma_1^2)\mathcal{N}(x|\mu_2,\sigma_2^2)\propto\mathcal{N}(x|\mu_n,\sigma_n^2),$$

with

$$\begin{aligned}&\mu_n=w_1\mu_1+w_2\mu_2,\quad w_1=\frac{\frac{1}{\sigma_1^2}}{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}}\text{ and }w_2=\frac{\frac{1}{\sigma_2^2}}{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}}\\ &\sigma_n^2=\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right)^{-1}.\end{aligned}$$