STA414/2104: Midterm Solutions

Section L5101

February 13, 2018

MC: F A D B B D B A A C A D A D B

1. (a)

$$p(\beta|x,a,b) = \frac{f(x|\beta)p(\beta|a,b)}{\int f(x|\beta)p(\beta|a,b)d\beta}$$

Just writing the correct answer gets full marks. If the students made a minor mistake, give 1 out of 2. Only give zero if the correct answer doesn't appear anywhere.

(b) Use the fact that since the prior is a distribution, it must integrate to 1, to obtain

$$\int \beta^{a-1}\exp(-b\beta)d\beta = \frac{\Gamma(a)}{b^a}$$

Therefore,

$$\begin{aligned}
&\int \frac{b^a}{\Gamma(a)}\beta^{a-1}\exp(-b\beta) \times \beta\exp(-x\beta)d\beta \\
&= \int \frac{b^a}{\Gamma(a)}\beta^{(a+1)-1}\exp\big(-(b+x)\beta\big)d\beta \\
&= \frac{b^a}{\Gamma(a)} \times \frac{\Gamma(a+1)}{(b+x)^{a+1}}
\end{aligned}$$

Note that since $\Gamma(a+1)/\Gamma(a) = a$, students may also have written

$$\frac{b^a a}{(b+x)^{a+1}}$$

which is also correct.

Minor calculation errors should lose 1 mark if the procedure was correct. If students tried to evaluate the integral by any other method other than using the fact that the prior integrates to 1, they either get full marks (if they somehow got it) or no marks, if they didn't get it.

(c) The posterior distribution is

$$\begin{aligned}
p(\beta|x,a,b) &= \frac{f(x|\beta)p(\beta|a,b)}{\int f(x|\beta)p(\beta|a,b)d\beta} \\
&= \frac{\frac{b^a}{\Gamma(a)}\beta^{(a+1)-1}\exp\big(-(b+x)\beta\big)}{\int \frac{b^a}{\Gamma(a)}\beta^{(a+1)-1}\exp\big(-(b+x)\beta\big)d\beta} \\
&= \frac{\frac{b^a}{\Gamma(a)}\beta^{(a+1)-1}\exp\big(-(b+x)\beta\big)}{\frac{b^a}{\Gamma(a)} \times \frac{\Gamma(a+1)}{(b+x)^{a+1}}} \\
&= \frac{(b+x)^{a+1}}{\Gamma(a+1)}\beta^{(a+1)-1}\exp\big(-(b+x)\beta\big)
\end{aligned}$$

which is recognized as the density of a $Gamma(a_1, b_1)$ distribution with

$$a_1 = a + 1$$

$$b_1 = b + x$$

If errors from previous parts cascaded down, you can give part marks if their procedure was correct. If they gave an unnormalized posterior (without dividing by the normalizing constant), take off 2 marks. If they showed it was gamma, gave the correct expression, but then didn't explicitly state a_1 and b_1 like the question asked, take off 1 mark.

(d) 2 marks for the correct formula for the predictive distribution

$$p(x^*|a, b) = \int f(x^*|\beta)p(\beta|x, a, b)d\beta$$

but don't give the mark if they plugged x^* into the posterior. The posterior depends on the original datapoint x; the new likelihood is for the new datapoint x^*. We have

$$\begin{aligned} p(x^*|a, b) &= \int f(x^*|\beta)p(\beta|x, a, b)d\beta \\ &= \frac{(b+x)^{(a+1)}}{\Gamma(a+1)} \int \beta^{(a+2)-1} e^{-(x+b+x^*)\beta} d\beta \\ &= \frac{(b+x)^{(a+1)}}{\Gamma(a+1)} \times \frac{\Gamma(a+2)}{(x+b+x^*)^{(a+2)}} \end{aligned}$$

2 marks for evaluating the correct answer. Take off 1 mark if they made a minor error, such as integrating with respect to x instead of β, or leaving the answer in integral form.

2. (a) The likelihood is

$$L(\lambda) = \prod_{n=1}^{N} P(Y_n = y_n) = \left(\prod_{n=1}^{N} y_n!\right) \times e^{-N\lambda} \times \lambda^{\left(\sum_{n=1}^{N} y_i\right)}$$

If they just wrote it as a product but didn't simplify, give them full marks as long as it's correct.

(b) In exponential family form,

$$L(\eta) = \left(\prod_{n=1}^{N} y_n!\right) \times e^{-Ne^{\eta}} \times \exp\left(\eta \sum_{n=1}^{N} y_i\right)$$

$$h(y_n) = y_n!$$

$$g(\eta) = e^{-e^{\eta}}$$

If they didn't write anything in terms of η, take off 2 marks.

(c) The MLE satisfies

$$-N\frac{\partial \log g(\eta)}{\partial \eta} = \sum_{n=1}^{N} u(y_n)$$

In this example,

$$-N\frac{\partial \log g(\eta)}{\partial \eta} = Ne^{\eta} = N\lambda = \sum_{n=1}^{N} y_i$$

$$\implies \hat{\lambda} = \bar{y}$$

If they did it manually (without using the exponential family identities), they get full marks.

(d) Now the η_n are different for each observation. Taking the product of the densities gives

$$\begin{aligned} L(\mathbf{w}) &= \prod_{n=1}^{N} h(y_n) \times e^{-\sum_{n=1}^{N} e^{\eta_n}} \times \exp\left(\sum_{n=1}^{N} \eta_n y_n\right) \\ &= \prod_{n=1}^{N} h(y_n) \times e^{-\sum_{n=1}^{N} e^{\mathbf{x}_n' \mathbf{w}}} \times \exp\left(\sum_{n=1}^{N} \mathbf{x}_n' \mathbf{w} y_n\right) \end{aligned}$$

(e) The MLE identity remains the same. We therefore have

$$\sum_{n=1}^{N} e^{\eta_n} \mathbf{x}_n = \sum_{n=1}^{N} \mathbf{x}_n y_n$$

3. (a) The k-Nearest-Neighbours algorithm as stated in lecture is

- For each $\mathbf{x}$ to be predicted,
  - Find the k closest points to $\mathbf{x}$ in Euclidean distance
  - Output the average target value for these points If students said the classification for each point is the most common classification of the neighbours, then take off 1 mark; this is not the fully general statement of the algorithm.

Students may have stated it mathematically, which is fine

(a) Define

$$N_k(\mathbf{x}) = \{\text{k closest points to x}\}$$

Then the k-Nearest-Neighbours algorithm is a linear smoother with

$$k(\mathbf{x}_n, \mathbf{x}) = \begin{cases} \frac{1}{k} \text{ if } x_n \in N_k(\mathbf{x}) \\ 0 \text{ else} \end{cases}$$

(b) The 1-NN approach. The data is well separated. The 100-NN approach would just classify all points as being halfway between classes. Give full marks if they just stated the answer. If they gave a wrong explanation, take off one mark.