Universidad de Oviedo

Modelos bayesianos. Mayo de 2024

Ejercicio 1 (2 puntos). En un estudio sobre preferencias de viaje se ha investigado la relación entre el tipo de destino (urbano, rural), el transporte (coche, tren, avión), la frecuencia de viajes realizados al año (mucha, poca) y la satisfacción con las experiencias del viaje (alta, baja). En dicho análisis se han llegado a estas conclusiones:
Pr(Transporte = avión) = 0.5,
Pr(Transporte = coche) = 0.3,
Pr(Transporte = tren) = 0.2,
Pr(Frecuencia = mucha) = 0.3,
Pr(Frecuencia = poca) = 0.7,
Pr(Destino = urbano | Transporte = avión, Frecuencia = mucha) = 0.8,
Pr(Destino = urbano | Transporte = coche, Frecuencia = mucha) = 0.6,
Pr(Destino = urbano | Transporte = tren, Frecuencia = mucha) = 0.4,
Pr(Destino = urbano | Transporte = avión, Frecuencia = poca) = 0.4,
Pr(Destino = urbano | Transporte = coche, Frecuencia = poca) = 0.2,
Pr(Destino = urbano | Transporte = tren, Frecuencia = poca) = 0.1,
Pr(Satisfacción = baja | Destino = urbano) = 0.2,
Pr(Satisfacción = baja | Destino = rural) = 0.4.

Justifique adecuadamente las respuestas.

1. ¿Son independientes el transporte y la frecuencia de viaje?
2. Dibuje un grafo que refleje la información aportada en el enunciado.
3. ¿Es condicionalmente independiente la satisfacción del viaje del medio de transporte dado el destino?
4. Supóngase que Pr(Destino = rural) = 0.606, Pr(Destino = rural | Transporte = tren)=0.81, Pr(Destino = rural, Transporte = tren)=0.162 y Pr(Satisfacción = baja) = 0.3112, ¿cuánto vale la probabilidad de que la satisfacción sea baja si se ha viajado en tren a un destino rural? (a) 0.4 (b) 0.6 (c) 0.52.

5. Con la información del apartado anterior y suponiendo que la respuesta correcta era la (a), ¿cuál es la probabilidad de ir a un destino rural cuando se ha viajado en tren y la satisfacción es baja? (a) 0.895 (b) 0.504 (c) 0.496.

Ejercicio 2.a (1 punto). Demuestre que el estimador más verosímil y el estimador máximo a posteriori coinciden cuando la distribución a priori es no informativa.

Ejercicio 2.b (1 punto). Reflexione sobre el siguiente texto y proponga argumentos a favor o en contra sobre las conclusiones que obtiene.

> «*Un ejemplo fácil de entender*. Tenemos una moneda y queremos estimar la probabilidad de cara con $n = 1$ [lanzamientos] (para que todo se vea más claro). Se tira la moneda y sale cara. Usando MLE [estimador máximo verosímil], el estimador de p es 100 %. Usando el modelo bayesiano de libro (con priori Beta(1, 1)), la posteriori sería una Beta(2, 1), que tiene la media en 2/3. Si para esa moneda $p = 0.5$, efectivamente, el modelo bayesiando sobreajusta y, efectivamente también, el MLE lo hace mucho más.
>
> »*Pero sobreajustar no es siempre necesariamente malo*. Imaginemos que tenemos un modelo genérico para las campañas publicitarias de los productos de una empresa, con sus prioris también genéricas. Ahora queremos construir un modelo para el producto A. Para ello, ajustamos el modelo usando una muestra de ventas de dicho producto.
>
> »Lo que hacemos al ajustar ese modelo es sobreajustar el modelo inicial que sirve para cualquier producto y adaptarlo a las peculiaridades específicas del producto A. [...] ilustra cómo, en el fondo, al aplicar modelos bayesianos buscamos la adaptación de modelos genéricos a casos más concretos. Y, en cierto modo, eso es sobreajustar.» Carlos

Gil Bellosta, *Los modelos bayesianos, ¿condenados a sobreajustar?*, 2024, https://www.datanalytics.com/.

Ejercicio 3 (2 puntos). Sea $\theta \sim Uniforme(0,1)$, $Y \sim Binomial(57,\theta)$, y la evidencia obtenida es $y = 19$. Se desea emplear el algoritmo de Metropolis para simular la distribución a posteriori. Justifique adecuadamente las respuestas.

1. ¿Cuáles son la distribución a priori, la función de verosimilitud y la distribución a posteriori?
2. ¿Cuál de las siguientes funciones de enlace utilizaría? (a) $x = \log\left(\frac{\theta}{1-\theta}\right)$ (b) $x = \log(\theta)$ (c) $x = \theta$.
3. ¿Es necesario emplear una función de salto para procesar este algoritmo? En caso afirmativo proponga una función de salto.
4. Sea s una iteración arbitraria del algoritmo. Supongamos que $\theta^{(s-1)} = 0.8$. Escriba un pseudocódigo del algoritmo Metropolis para determinar un valor candidato $\theta^{(*)}$ para esta iteración.

Ejercicio 4 (2 puntos). Sea $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$ un conjunto de variables aleatorias tales que siguen la misma distribución, $\beta_0, \beta_1, \ldots, \beta_p \sim Laplace(0,1)$ y son independientes. Se define la variable aleatoria Y de la siguiente forma

$$p(y_i \mid \boldsymbol{\beta}) \propto \exp\left(-\frac{1}{2}(y_i - \boldsymbol{x}_i^T \boldsymbol{\beta})^2\right),$$

con $\boldsymbol{x}_i = (x_{i,0}, x_{i,1}, \ldots, x_{i,p})$ vectores de datos fijos para $i = 1, \ldots, n$ e y_i independientes entre sí. Justifique adecuadamente las respuestas.

1. Explicite la distribución conjunta de $\boldsymbol{\beta}$.
2. Determine una expresión proporcional a la distribución de

$$p(y_1, y_2, \ldots, y_n \mid \boldsymbol{\beta}).$$

3. Determine una expresión proporcional a $p(\boldsymbol{\beta} \mid y_1, y_2, \ldots, y_n)$.
4. El estimador máximo a posteriori de $\boldsymbol{\beta}$, ¿sirve como solución

$\hat{\boldsymbol{\beta}}_{\text{lasso}}$ a una regresión de tipo lasso?

$$\hat{\boldsymbol{\beta}}_{\text{lasso}} = \arg\min_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^{n} (y_i - \boldsymbol{x}_i^T \boldsymbol{\beta})^2 + \lambda \sum_{j=0}^{p} |\beta_j| \right\}.$$

Ejercicio 5 (2 puntos). Se quiere llevar a cabo una evaluación de riesgos en el desarrollo de una herramienta de software encargada por una determinada empresa. Para llevar a cabo esa evaluación se modeliza el problema a través de una red bayesiana considerando las siguientes variables: «Experiencia del equipo de desarrollo» (E), «Complejidad del proyecto» (CP), «Calidad del código» (CC), «Cambios en los requisitos durante el desarrollo» (C), «Retraso en la entrega del proyecto» (R) y «Evaluación global del proyecto» (EG).

1. Proporcione un conjunto de valores razonables para cada una de las variables anteriores.
2. Realice un diagrama a través de un grafo acíclico dirigido representando las posibles implicaciones causales que puedan darse en el grafo explicando brevemente porqué considera ésas.
3. Proponga unas tablas de probabilidad condicionada que sean coherentes con el grafo del apartado anterior.
4. ¿Cuál es, en la red bayesiana propuesta, la cobertura de Markov de la variable «Evaluación global del proyecto»?
5. Si se considera el conjunto de variables X = {CP, CC}, ¿están d-separadas las variables E y EG dado X?
6. Determine la distribución marginal de probabilidades de la variable CC.