**Esenciales bayesianos con R**

**8 Análisis de imagen**

**Mapa vial**

Este capítulo final cubre el análisis de imágenes pixeladas a través de modelos de campo aleatorio de Markov, hacia la detección de patrones y la corrección de imágenes. Comenzamos con el análisis estadístico de los campos aleatorios de Markov, que son extensiones de las cadenas de Markov al dominio espacial, ya que son fundamentales en este capítulo. Esta es también la oportunidad perfecta para cubrir el método ABC, ya que estos modelos no permiten una probabilidad de forma cerrada. El análisis de imágenes ha sido un área muy activa tanto para las estadísticas bayesianas como para los métodos computacionales en los últimos 30 años, por lo que creemos que merece un capítulo propio por sus características específicas.

**8.1 El análisis de imágenes como problema estadístico**

Si pensamos en una imagen de computadora como una colección (grande) de píxeles de colores dispuestos en una cuadrícula, ¡no parece haber ninguna aleatoriedad involucrada ni necesidad de análisis estadístico! No obstante, el análisis de imágenes visto como un análisis estadístico es un campo próspero que vio el surgimiento de varios avances estadísticos importantes, incluido, por ejemplo, el muestreador de Gibbs. (Además, este campo ha adoptado predominantemente una perspectiva bayesiana tanto porque esto era algo natural de hacer como porque el poder analítico de este enfoque era mayor que con otros métodos). La razón de esta aparente paradoja es que, mientras que los píxeles suelen ser deterministas objetos, la complejidad y el tamaño de las imágenes requieren que uno represente esos píxeles como la salida aleatoria de una distribución gobernada por un objeto de dimensión mucho más pequeña. Por ejemplo, este es el caso de la visión por computadora, donde los objetos específicos deben extraerse de un fondo mucho más rico (o más ruidoso).

En este espíritu de extraer información de una enorme estructura dimensional, construimos en la Secta. 8.2 una familia específica de distribuciones inspiradas en la física de partículas, el modelo de Potts, para estructurar imágenes y otras estructuras espaciales en términos de homogeneidad local. Desafortunadamente, esta es una sección principalmente teórica con muy pocas ilustraciones. En la secc. 8.3, abordamos el problema fundamental de manejar la constante de normalización faltante en estos modelos mediante la introducción de una nueva técnica computacional llamada ABC que opera sobre verosimilitudes intratables (con la penalidad de producir una respuesta aproximada). En la secc. 8.4, imponemos una fuerte dimensión espacial al prior asociado a una imagen con el fin de reunir estructuras homogéneas a partir de una imagen compleja o borrosa.

**8.2 Dependencia espacial**

**8.2.1 Rejillas y Lattices**

Una imagen (en el sentido de una imagen generada por computadora) es un caso especial de una celosía, en el sentido de que es un objeto aleatorio cuyos elementos están indexados por la ubicación de los píxeles y, por lo tanto, están relacionados por la proximidad geográfica de esas ubicaciones. En general, una celosía es un objeto matemático multidimensional en el que se puede definir una relación de vecindad.

A pesar de que el análisis original de modelos de celosía de Besag (1974) se centró en la ecología vegetal y los experimentos agrícolas, la relación de vecindad solo se limita a ser una relación simétrica y no necesariamente tiene una conexión con una proximidad geográfica, ni con una imagen. Por ejemplo, la relación puede describir interacciones sociales entre tribus amazónicas o palabras en un manuscrito que comparten una raíz lingüística. (La relación de vecindad entre dos puntos de la celosía generalmente se traduce en términos estadísticos en una dependencia probabilística entre esos puntos.) La celosía asociada con una imagen es una matriz regular n × m hecha de (i, j) 's (1 ≤ i ≤ n, 1 ≤ j ≤ m), cuyos vecinos más cercanos (pero no necesariamente solo) están formados por las cuatro entradas (i, j - 1), (i, j + 1), (i - 1, j) y (i + 1, j). Para describir adecuadamente una estructura de dependencia en imágenes o en otros objetos espaciales indexados por una celosía, necesitamos expandir la noción de cadena de Markov en esas estructuras. Dado que una celosía es un objeto multidimensional —a diferencia de la línea unidimensional correspondiente a los tiempos de observación de la cadena de Markov—, un primer requisito para la generalización es definir una estructura de vecindad adecuada.

Para ilustrar esta noción, consideramos un pequeño conjunto de datos que muestra la presencia de juncos copetudos en una parte de un humedal. Este conjunto de datos, llamado Laichedata, es simplemente una matriz de ceros y unos de 25 × 25. La celosía correspondiente es la matriz de 25 × 25 (Fig. 8.1).

Dada una red I de sitios i ∈ I en un mapa o de píxeles en una imagen, 3 una relación de vecindad en I se denota por ∼, i ∼ j significa que i y j son vecinos. Si asociamos una distribución de probabilidad en un vector x indexado por la red, x = (xi) i∈I, con esta relación, lo que significa que dos componentes xi y xj están correlacionados si los sitios iyj son vecinos, un requisito fundamental para la existencia de esta distribución es que la relación de vecindad es simétrica (Cressie, 1993): si i es un vecino de j (escrito como i ∼ j), entonces j es un vecino de i. (Por convención, i no es vecino de sí mismo.) La figura 8.2 ilustra esta noción para tres tipos de vecindarios en una cuadrícula regular. Por ejemplo, Laichedata podría asociarse con un vecindario noroeste-sureste para tener en cuenta los vientos dominantes: una entrada (i, j) tendría como vecinos (i - 1, j - 1) y (i + 1, j + 1).

**8.2.3. El modelo de Ising**

Si los píxeles de la imagen x bajo estudio sólo pueden tomar dos colores (blanco y negro, digamos, como en la figura 8.1), x es binario. Normalmente nos referimos a cada píxel xi como primer plano si xi = 1 (negro) y fondo si xi = 0 (blanco). La distribución condicional de un píxel es entonces Bernoulli, con el parámetro de probabilidad correspondiente dependiendo de los otros píxeles. Un paso de simplificación es asumir que es una función del número de píxeles vecinos negros, utilizando, por ejemplo, un enlace logit como (j = 0, 1)

donde ni, j = 􏰂l∈n (i) Ixl = j es el número de vecinos de xi con color j. El modelo de Ising se define luego a través de estos condicionales completos

y la distribución conjunta por lo tanto satisface

donde la suma se toma sobre todos los pares (i, j) de vecinos (ejercicio 8.17). Al inferir sobre β y, por tanto, simular la distribución posterior β, nos enfrentaremos a un obstáculo importante, a saber, que la constante de normalización de (8.4), Z (β), es intratable, excepto para las redes muy pequeñas I, mientras que depende de β. Por lo tanto, no se puede calcular la función de verosimilitud. Introduciremos en la Secta. 8.3 una técnica computacional llamada ABC que está destinada a combatir este mismo problema.

En esta etapa inicial, sin embargo, consideramos que β es conocido y nos enfocamos en la simulación de x en preparación para la inferencia tanto de β como de x dada una versión ruidosa de la imagen, y, como se presenta en la Sección. 8.4.

El enigma computacional de los modelos de Ising es más profundo ya que, debido a la intrincada estructura de correlación del modelo de Ising, no es posible una simulación directa de x, espere en casos muy específicos. Frente a esta dificultad, la comunidad de imágenes desarrolló muy pronto herramientas computacionales que finalmente llevaron en 1984 a la propuesta del gibbs sampler.

La especificación de los campos aleatorios de Markov y, en particular, del modelo de Ising implica que las distribuciones condicionales completas de esos modelos están disponibles en forma cerrada. La estructura local de los campos aleatorios de Markov proporciona una actualización inmediata sitio por sitio para el muestreador de Gibbs:

En esta implementación, el orden de las actualizaciones de los píxeles de I es aleatorio para poder superar posibles cuellos de botella en la exploración de la distribución, aunque esta no es una condición necesaria para que el algoritmo converja. De hecho, cuando se consideran dos píxeles x1 y x2 que están separados por m píxeles, la influencia de un cambio en x1 no se siente en x2 antes de al menos m iteraciones del muestreador Gibbs básico. Por supuesto, si m es grande, la dependencia entre x1 y x2 es bastante moderada, pero esta lenta propagación de cambios es indicativa de una mezcla lenta en la cadena de Markov. Por ejemplo, ver un cambio de color en una región homogénea relativamente grande es un evento de muy baja probabilidad, aunque la distribución de los colores es intercambiable.

Junto con la dinámica lenta inducida por la actualización de un solo sitio, podemos señalar otra ineficiencia de este algoritmo, a saber, que muchas actualizaciones no modificarán el valor actual de x simplemente porque el nuevo valor de xl es igual a su valor anterior. Sin embargo, es sencillo modificar el algoritmo para que solo proponga cambios de valores. La actualización de cada píxel l es entonces un paso de Metropolis-Hastings con probabilidad de aceptación

**8.2.4 El modelo de Potts**

La generalización del modelo de Ising a los casos en que la imagen tiene más de dos colores dice G, es sencilla. Si ni, g denota el número de vecinos de i ∈ I con color g (1 ≤ g ≤ G), es decir,

la distribución condicional completa de xi se elige como

Esta elección corresponde a un modelo de probabilidad conjunta (verdadero), el modelo de Potts, cuya densidad viene dada por (Ejercicio 8.6)

Este modelo es una generalización clara del modelo de Ising y tiene el mismo inconveniente, a saber, que la constante de normalización de esta densidad, que es una función de β, no está disponible en forma cerrada y, por lo tanto, dificulta la inferencia y el cálculo de la verosimilitud. función.

Una vez más, nos enfrentamos al obstáculo de que, al simular x de un modelo de Potts con un β grande, el muestreador de Gibbs de un solo sitio puede ser bastante lento. Hay alternativas más eficientes disponibles, incluido el algoritmo de Swendsen-Wang (ejercicio 8.7). Por ejemplo, el algoritmo 8.17 a continuación es nuevamente un algoritmo de Metropolis-Hastings que fuerza los movimientos en los valores actuales. Tenga en cuenta la característica especial de que, si bien esta propuesta de Metropolis-Hastings no es una caminata aleatoria, usar en su lugar una propuesta uniforme sobre los otros valores posibles G− 1 aún conduce a una probabilidad de aceptación que es igual a la relación de las densidades objetivo.

La figura 8.4 ilustra el resultado de una simulación usando el algoritmo 8.17 en una situación donde hay G = 4 colores, usando la siguiente función R para la simulación. (El uso de un solo vector de índices para filas y columnas es un truco de programación que elimina un bucle en el código y, por lo tanto, ahorra una cantidad considerable de tiempo de cálculo. Esto también permite una verdadera distribución uniforme en la muestra. Tenga en cuenta la llamada a la congruencia operadores %% para módulo y% /% para división de enteros) Señalamos la influencia reforzada de los β grandes en la Fig. 8.4: no solo es mayor la homogeneidad, sino que también hay una mayor diferenciación en los colores.10 Destacamos que, mientras que β en la figura 8.4 varía sobre los mismos valores que en la figura 8.3, las β no son directamente comparables ya que el mayor número de clases en el modelo de Potts induce un valor menor de las ni, g para la estructura de vecindad.

**8.4. Segmentación de imagen**

En esta sección, todavía consideramos las imágenes como objetos estadísticos, pero ahora son "ruidosas" en el sentido de que el color o el nivel de gris de un píxel no se observa exactamente, pero con alguna perturbación (a veces llamado desenfoque como en las imágenes de satélite). El propósito de la segmentación de imágenes es agrupar píxeles en clases homogéneas sin supervisión o definición preliminar de esas clases, basándose únicamente en la coherencia espacial de la estructura.

Esta estructura subyacente de los píxeles "verdaderos" se denota por x, mientras que la imagen observada se denota por y. Ambos objetos x e y son matrices, con cada entrada de x tomando un número finito de valores y cada entrada de y tomando valores reales (por conveniencia de modelado en lugar de restricciones de realidad). Por lo tanto, estamos interesados en la distribución posterior de x dada y proporcionada por el teorema de Bayes, π (x | y) ∝ f (y | x) π (x). En esta distribución posterior, la probabilidad, f (y | x), describe el vínculo entre la imagen observada y la clasificación subyacente; es decir, da la distribución del ruido, mientras que el π (x) anterior codifica creencias sobre las propiedades (posibles o deseadas) de la imagen subyacente. Aunque, como en otros capítulos, no podemos proporcionar la historia completa de la segmentación de imágenes bayesianas, se puede encontrar un excelente tutorial sobre el procesamiento de imágenes bayesianas basado en un curso de la escuela de verano en Hurn et al. (2003).

Como se indicó anteriormente, una motivación adecuada para la segmentación de imágenes es el procesamiento de satélites, ya que las imágenes captadas por satélites a menudo se ven borrosas, ya sea debido a inexactitudes en los instrumentos o transmisión o debido a nubes o cobertura de vegetación entre el satélite y el área de interés.

Con el modelo que se está introduciendo, pasamos al tema central, a saber, cómo hacer inferencias sobre la imagen "verdadera", x, dada una imagen ruidosa observada, y. El anterior en x es un modelo de Potts con categorías G,

donde Z (β) es la constante de normalización (intratable, véase la sección 8.3) del modelo de Potts. Dado x, asumimos que las observaciones en y son variables aleatorias normales independientes,

Este modelo no es exacto ya que las yi son niveles de gris enteros que varían entre 0 y 255, pero es más fácil de manejar que una distribución parametrizada en {0,. . . , 255}. Esta configuración recuerda claramente14 a los modelos de Markov mezclados y ocultos de los capítulos 6 y 7 en el sentido de que una estructura de Markov, el campo aleatorio de Markov, solo se observa a través de variables aleatorias indexadas por los estados.

En este problema, los parámetros β, σ2, μ1, ..., μG se suelen considerar parámetros molestos, punto de vista que justifica el uso de previos uniformes como

el último anterior corresponde a un previo uniforme en log σ.

El límite superior de β se analizó en la sección anterior. El orden de los μg no es necesario, estrictamente hablando, pero evita el fenómeno de cambio de etiqueta discutido en la Sección 6.5. (La alternativa es usar el mismo uniforme antes en todos los μg y luego reordenarlos una vez que se realiza la simulación MCMC. Si bien esto puede evitar comportamientos de convergencia lenta en algunos casos, esta estrategia también implica una contabilidad más complicada y mayores requisitos de almacenamiento. . En el caso de imágenes grandes, simplemente no se puede considerar.)