MCMC для dummies

Винни-Пух

3 марта 2017 г.

Предисловие

План:

1. Энтропия/Джини

Распределения максимизирующие энтропию? что-то про ROC кривые до кучи?

2. Одиноко стоящий дуб

Типичное заданичие: Вырастить дерево согласно такому-то критерию. Сюда борьбу с NA. Сюда же регуляризацию? Или отдельно?

3. Логит-модель

Логистическое распределение? Перевод y=0/1 в y=-1/1. Максимум правдоподобия в минимум штрафа? Предельные эффекты?

4. Мини-мими-лес

Типичное: Два-три дерева. По ним построить прогноз/оценить важность переменных. Что еще?

5. Регуляризация.

Общая идея. Парадокс James-Stein. Для среднего, для регрессии, для дерева. L1 и L2.

6. Про кросс-валидацию?

Как это делать руками? Какие тут теоретические задачи?

Упр: Дано одно-два-три дерева. И 5 наблюдений. Посчитать кросс-валидационную ошибку.

Упр: На наборе данных в 5 наблюдений подобрать параметр жесткости с помощью кросс-валидации.

7. Несколько практических упражнений.

Упр: сделайте с дефолтными параметрами и ответьте на все подробности про алгоритм
тут решения в python/R.

Упр: Нарисуйте дерево номер 5.

Из теории:

- определения
- табличка с параметрами xgboost, rforest
- несколько практик подбора параметров

```r
library("knitr") # грамотное программирование
library("tikzDevice") # сохранение графиков в формате tikz

library("tidyverse") # Хэдли на нашей стороне
library("xtable")

theme_set(theme_bw()) # чёрно-белая тема для графиков
```

Глава 1

Неразобранные :)

http://www.stat.columbia.edu/~gelman/book/solutions2.pdf

1.1. Марковские цепи

1.1 Шахматный конь начинает в клетке A1. Каждый свой ход он выбирает равновероятно из возможных. Какова вероятность того, что через много-много ходов он окажется в клетке H8? Сколько в среднем длится путь от клетки A1 до клетки A1?

1.2. Ручные задачи

1.2 Случайные величины X_i независимы и одинаково распределены с табличкой

X	1	2	6
$\mathbb{P}()$	β	2β	$1-3\beta$

Известно, что $X_1 = 1$, $X_2 = 2$, $X_3 = 2$, $X_4 = 4$.

1. Найдите оценку $\hat{\beta}$ методом моментов.
2. Найдите оценку $\hat{\beta}$ методом максимального правдоподобия.

Априорно известно, что β равномерно на отрезке $[0; 1/3]$.

3. Найдите апостериорную условную функцию плотности β с учётом полученных наблюдений.
4. Сгенерируйте выборку из апостериорного распределения с помощью STAN.

Априорно известно, что β имеет функцию плотности $f(t) = 18t$ на отрезке $[0; 1/3]$.

5. Найдите апостериорную условную функцию плотности β с учётом полученных наблюдений.
6. Сгенерируйте выборку из апостериорного распределения с помощью STAN.

1.3 Маша прячется от Медведей в точке m на числовой прямой. Есть несколько Медведей, каждый из которых обнюхивает всю числовую прямую в поисках Маши. Медведю номер i кажется, что Машей сильней всего пахнет в точке y_i. Естественно, Медведи могут ошибаться, например, у них может быть заложен нос, поэтому $y_i|m \sim \mathcal{N}(m, 2^2)$. При фиксированном m величины y_i независимы.

Известно, что $y_1 = 0.5$, $y_2 = -1$.

1. Найдите оценку $\hat{m}$ методом моментов.
2. Найдите оценку $\hat{m}$ методом максимального правдоподобия.

Априорно известно, что m нормально $\mathcal{N}(1, 4^2)$:

3. Найдите апостериорную условную функцию плотности m с учётом полученных наблюдений.
4. Сгенерируйте выборку из апостериорного распределения с помощью STAN.

Априорно известно, что m равномерно на отрезке $[0; 10]$:

5. Найдите апостериорную условную функцию плотности m с учётом полученных наблюдений.
6. Сгенерируйте выборку из апостериорного распределения с помощью STAN.

1.4 дискретная переменная — непрерывные наблюдения

1.5 дискретная переменная — дискретные наблюдения

1.3. toy stan

1.6 Напишите на языке STAN модель парной регрессии. А именно, априорно предполагается, что

$$\begin{cases} \beta_1 \sim \mathcal{N}(0, 10^2) \\ \beta_2 \sim \mathcal{N}(1, 10^2) \\ \sigma^2 \sim IG(rate = 3, shape = 4) \end{cases}$$

Наблюдаемые y_i порождаются согласно уравнению

$$y_i = \beta_1 + \beta_2 x_i + u_i,\ u_i \sim \mathcal{N}(0; \sigma^2)$$

На вход модели должны подаваться количество наблюдений N, значения зависимой переменной y_1, ..., y_N и значения независимой переменной x_1, ..., x_N.

1.7 Напишите на языке STAN модель процесса AR(1). А именно, априорно предполагается, что

$$\begin{cases} \beta_1 \sim \mathcal{N}(0, 10^2) \\ \beta_2 \sim \mathcal{N}(1, 10^2) \\ \sigma^2 \sim IG(rate = 3, shape = 4) \end{cases}$$

Наблюдаемые y_i порождаются согласно уравнению

$$y_i = \beta_1 + \beta_2 y_{i-1} + u_i,\ u_i \sim \mathcal{N}(0; \sigma^2)$$

На вход модели должны подаваться количество наблюдений N, значения зависимой переменной y_1, ..., y_N. На выходе должна получаться выборка из апостериорного распределения β_1, β_2 и σ^2.

1.8 Напишите на языке STAN байесовский тест для сравнения средних при неизвестных дисперсиях и малом количестве наблюдений.

$$\begin{cases} \mu_A \sim \mathcal{N}(0, 100^2) \\ \mu_B \sim \mathcal{N}(0, 100^2) \\ \sigma_A^2 \sim IG(rate = 3, shape = 4) \\ \sigma_B^2 \sim IG(rate = 3, shape = 4) \end{cases}$$

При фиксированных параметрах распределения условно независимы и $y_i^A | \mu_A, \sigma_A^2 \sim \mathcal{N}(\mu_A, \sigma_A^2)$, $y_i^B | \mu_B, \sigma_B^2 \sim \mathcal{N}(\mu_B, \sigma_B^2)$.
Даны конкретные наблюдения: $y_1^A = 2.7$, $y_2^A = 3.8$, $y_3^A = 4.1$, $y_1^B = 4.2$, $y_2^A = 3.8$, $y_3^A = 3.1$.

1. Постройте выборку из апостериорного распределения.
2. Оцените апостериорную вероятность того, что μ_B более чем на единицу превосходит μ_A.

1.4. stan - integrate out

1.9 Рассмотрим данные по количеству происшествий на английских шахтах. Предположим, что до момента b включительно количество происшествий за год имеет пуассоновское распределение с параметром λ_1, а после момента b — пуассоновское с параметром λ_2.

Всего есть данные за T дней. Предположим, что априорно b равновероятно принимает значения от 1 до T, а $\lambda_i \sim IG(rate = 3, shape = 4)$.

Фактические данные по количеству происшествий за год рассчитайте исходя из дат происшествий по ссылке http://people.reed.edu/~jones/141/Coal.html.

С помощью STAN:

1. Постройте выборку из апостериорного распределения λ_1, λ_2, b.
2. Какой год наиболее вероятно был годом структурного сдвига?
3. Постройте прогноз распределения количества аварий на шахтах на следующий год.

1.5. Гиббс

1.10 Используя алгоритм Гиббса сгенерите выборку для двумерного нормального распределения $N(0, A)$, $A = \begin{pmatrix} 4 & -1 \\ -1 & 2 \end{pmatrix}$.

1.6. Метрополис-Гастингс

1.11 Используя алгоритм Метрополиса-Хастингса сгенерите выборку для биномиального распределения $Bin(n, p)$ из равновероятного на множестве $\{0, 1, 2, \ldots, n\}$:

1. Подробно опишите алгоритм.
2. Реализуйте алгоритм на каком-нибудь языке программирования.

1.12 Используя алгоритм Метрополиса-Хастингса сгенерите выборку для биномиального распределения $Bin(n, p)$ из симметричного случайного блуждания на $\mathbb{Z}$.

1. Подробно опишите алгоритм.
2. Реализуйте алгоритм на каком-нибудь языке программирования.

1.13 Используя алгоритм Метрополиса-Хастингса сгенерите выборку для геометрического распределения $Geom(p)$ из симметричного случайного блуждания на $\mathbb{Z}$.

1. Подробно опишите алгоритм.
2. Реализуйте алгоритм на каком-нибудь языке программирования.

1.14 Используя алгоритм Метрополиса-Хастингса сгенерите выборку для пуассоновского распределения $Pois(\lambda)$ из симметричного случайного блуждания на $\mathbb{Z}$.

1. Подробно опишите алгоритм.
2. Реализуйте алгоритм на каком-нибудь языке программирования.

1.15 Используя алгоритм Метрополиса-Хастингса сгенерите выборку для функции плотности $\pi(x) \sim \exp(-x^2)(3 + x^2 + \cos x)$ из нормального $N(0, 1)$. Из нормального $N(0, \sigma^2)$.

1. Подробно опишите алгоритм.
2. Реализуйте алгоритм на каком-нибудь языке программирования.

1.16 Используя алгоритм Метрополиса-Хастингса сгенерите выборку для функции плотности $\pi(x) \sim \exp(-x^2)(3 + x^2 + \cos x)$ из случайного блуждания $X_{t+1} = X_t + \varepsilon_t$, где $\varepsilon_t \sim N(0, 1)$. Вариант с $N(0, \sigma^2)$.

1. Подробно опишите алгоритм.
2. Реализуйте алгоритм на каком-нибудь языке программирования.

1.17 Используя алгоритм Метрополиса-Хастингса сгенерите выборку для стандартного нормального распределения $N(0, 1)$ из случайного блуждания $X_{t+1} = X_t + \varepsilon_t$, где $\varepsilon_t \sim U[-1, 1]$.

1. Подробно опишите алгоритм.
2. Реализуйте алгоритм на каком-нибудь языке программирования.

1.18 Используя алгоритм Метрополиса-Хастингса сгенерите выборку для двумерного нормального распределения $N(0, A)$, $A = \begin{pmatrix} 4 & -1 \\ -1 & 2 \end{pmatrix}$ из случайного блуждания $X_{t+1,i} = X_{t,i} + \varepsilon_{t,i}$, где $\varepsilon_{t,i} \sim U[-1, 1]$.

1. Подробно опишите алгоритм.
2. Реализуйте алгоритм на каком-нибудь языке программирования.

1.19 Используя алгоритм Метрополиса-Хастингса сгенерите выборку для двумерного распределения с функцией плотности $p(x, y) = \exp(-4x^2 - 6y^2 + 2x - y + xy)$, $x > 0, y > 0$ из случайного блуждания $X_{t+1,i} = X_{t,i} + \varepsilon_{t,i}$, где $\varepsilon_{t,i} \sim U[-1, 1]$.

1. Подробно опишите алгоритм.
2. Реализуйте алгоритм на каком-нибудь языке программирования.

1.7. Регуляризация

1.20 Рассмотрим модель

$$y = X\beta + u,$$

где u_i независимы и $\mathcal{N}(0; \sigma^2)$.

Метод гребневой регрессии предполагает минимизацию функции

$$Q(\beta) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{k} \beta_j^2.$$

Рассмотрим байесовский подход к регрессии. Предположим, что априорное распределение имеет вид $\sigma^2 \sim InvGamma(a, b)$, $\beta_j | \sigma^2 \sim \mathcal{N}(0; c(\sigma^2))$.

При каких a, b и $c(\sigma^2)$ апостериорная мода $\hat{\beta}_{MAP}$ совпадёт с $\hat{\beta}_{Ridge}$?

1.21 Рассмотрим модель

$$y = X\beta + u,$$

где u_i независимы и $\mathcal{N}(0; \sigma^2)$.

Метод LASSO предполагает минимизацию функции

$$Q(\beta) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{k} |\beta_j|.$$

Рассмотрим байесовский подход к регрессии. Предположим, что априорное распределение имеет вид $\sigma^2 \sim InvGamma(a, b)$, $\beta_j|\sigma^2 \sim DoubleExp(c(\sigma^2))$.

При каких a, b и $c(\sigma^2)$ апостериорная мода $\hat{\beta}_{MAP}$ совпадёт с $\hat{\beta}_{LASSO}$?

1.22 Храбрый Охотник ловит Покемонов в случайном порядке. Вес i-го пойманного Покемона, y_i, имеет нормальное распределение $\mathcal{N}(\mu; \sigma^2)$. Параметры μ и σ неизвестны.

Храбрый охотник хочет оценить μ по формуле $\hat{\mu} = c \sum_{i=1}^{n} y_i$.

1. При каком c величина $\mathbb{E}((\hat{\mu} - \mu)^2)$ будет минимальна?
2. Возможно ли использовать на практике данное c?

1.8. Сопряжённые распределение

Надёргать: википедия, статья Айвазяна, примеры у Tsay

1.23 При известном параметре p, наблюдаемые $y_1, \ldots, y_n$ независимы и имеют распределение Бернулли, $y_i|p \sim Bernoulli(p)$. Априорно параметр p имеет бета-распределение, $p \sim beta(\alpha, \beta)$.

1. Найдите апостериорное распределение параметра p.
2. Найдите апостериорный прогнозный закон распределения y_{n+1}.

Глава 2

Решения и ответы к избранным задачам

1.1.

1.2.

1.3.

1.4.

1.5.

1.6.

1.7.

1.8.

1.9.

1.10.

1.11.

1.12.

1.13.

1.14.

1.15.

1.16.

1.17.

1.18.

1.19.

1.20.

1.21. При любых a и b, и $c(\sigma^2) = \lambda/2\sigma^2$

1.22. $c = \frac{1}{n+\sigma^2/\mu^2}$, нет, так как μ и σ^2 неизвестны.

1.23.

Список обозначений

Оглавление