

Диплом все дела

Егоров Евгений

May 2017

Содержание

1	Предсказание землетрясений	3
1.1	Постановка задачи	3
1.2	Прекурсоры	4
2	B-value прекурсор	6
2.1	Обзор	6
2.2	Распределение статистики	7
3	RTL прекурсор	9
3.1	Обзор	9
3.2	Определение критических значений	10
4	Новый прекурсор на основе b-value	11
4.1	Вероятностная модель	11
4.2	Вероятностный вывод	12
4.3	Ядровая версия	13
4.4	Байесовский прогноз	14
5	Траектории землетрясений как смесь регрессий	15
5.1	Теория вариационного вывода	15
5.2	Смесь регрессий	17
5.3	Отличие от других подходов	17
5.4	Вероятностный вывод	18
5.5	Иллюстративные примеры на искусственных данных	21
5.6	Кластеризация вдоль траектории	22
5.7	Использование для предсказания землетрясений	22
6	Предсказание землетресения	24
6.1	Агрегация различных прекурсоров	24
7	Результаты	26
7.1	Япония	26
7.2	Италия	27
7.3	Выводы	27

1 Предсказание землетрясений

1.1 Постановка задачи

Пусть t индексирует время. Мы наблюдаем временной ряд Y_t , где $Y_i = [\text{целевой событие в момент } t = i]$. Также мы наблюдаем признаки $X_{t;k}$, для каждого фиксированного $k = k'$, $X_{t;k'}$ многомерный временной ряд. Неформально, наша цель состоит в заблаговременном предсказании целевого события Y по признаковому описанию X . Иначе говоря: по наблюдаемой истории $(X_{\{t,x\}}; Y_t)$ до некоторого момента T , $t \in [0; T]$, необходимо поднять тревогу в окне $[T; T + \delta_1]$ при том, что целевой событие происходит в окне $[T + \delta_1; T + \delta_1 + \delta_2]$. Примерами таких задач может служить:

- Предсказание поломок в сложных технических системах. Тогда k индексирует датчики в такой системе и $X_{t;k}$ представляет из себя показания датчика.
- Предсказание природных катастроф. Тогда k может представлять или координаты $(lat; long)$ либо номер кластера.

В данной работе мы фокусируемся на предсказании землетрясений. Необходимо определить, что тогда является целевым событием. Существует несколько вариантов:

- Предсказание начала серии афтершоков, т.е. последовательности землетрясений после самого сильного (main shock). С такой задачей хорошо справляются различные ETAS модели.
- Предсказание сильных землетрясений на долгосрочном горизонте (годы) (модели seismic gap)
- Предсказание сильных землетрясений на среднесрочном горизонте (месяцы)

Мы займемся предсказанием сильных землетрясений на среднесрочном горизонте. Под сильными имеется ввиду землетрясения, с магнитудой выше $M_c = 5$. Перечислим сложности, с которыми связаны предсказания землетрясений:

- Очень не сбалансированная выборка

С 1990-2016 в Японии, всего 247204 землетрясений, распределение по магнитудам:

Магнитуда	Число землетрясений с большей магнитудой
5.5	759 (0.3%)
6.5	87 (0.035%)
7.5	7 (0.002%)
8	1

- Искусственные аномалии в каталоге из-за изменение сети сейсмо-станций
- Чем выше магнитуда ожидаемого распределения, тем шире окно между прекурсором и им (более года для землетрясение магнитудой больше 7)
- Quasi-Periodic, Poisson, Clustered? Есть свидетельства как против так и за каждой гипотезы
- Сложно валидировать методики, т.к. прогнозы могут быть отложены во времени [1]: "there is a 0.62 probability of a major, damaging [$M \geq 6.7$] earthquake striking the greater San Francisco Bay Region over the next 30 years (2002-2031)"

1.2 Прекурсоры

Базовая предпосылка: появление сильных землетрясений робастно описывается функционалами $F_k(t)$, каждый из которых описывает какой-либо статистический паттерн сейсмического поля и рассчитывается в скользящем окне $(t - s, t)$. Если аномальные значения такого функционала отвечают наступлению сильного землетрясения, то будем называть такой функционал прекурсором.

Прекурсоры строятся как статистики, зависящие от:

- изменения сейсмической активности

- изменение кластеризации землетрясений во времени или пространстве
- изменения зависимости частоты землетрясений от магнитуды

Возможные данные, на которых строят прекурсоры:

- Геомагнитное поле
- Концентрация радона
- Изменений уровня/температуры подземных вод
- Деформация земной поверхности
- Изменение в бассейнах нефти
- Микросейсмичность (высочастотные ряды)
- **Мы работаем с каталогом землетрясений:** (latitude, longitude, depth, magnitude, Year-Month-Day-Hour-Seconds), относительно низкочастотные

Далее рассмотрим популярные прекурсоры и предложим свой.

2 B-value прекурсор

2.1 Обзор

Прекурсор был введен в статье [2] и известен также как Gutenberg–Richter law.

$\log(N(> M)) = a - bM$, $N(> M)$ общее число землетрясений с магнитудой больше M

Изначально был обнаружен чисто как эмпирический факт, позднее возникли различные обоснования в геофизике, обзор которых дан в статье [3]. Однако, до сих пор нет подтверждения или обоснования, что пространственные и временные изменения в b-value действительно являются предвестниками сильных землетрясений. Некоторый эмпирический опыт можно обобщить как [4, 5, 6]:

- отклонение b-value от среднего уровня по пространству отвечает повышению уровня напряжения между тектоническими плитами
- связь обратная, резкое уменьшение b-value отвечает сильному напряжению
- отклонение b-value от среднего уровня по времени отвечает возможному сильному землетрясению
- здесь связь прямая: сильным землетрясением предшествует положительно отклонение b-value от среднего уровня по времени

Перечислим также недостатки, которые бы хотелось отметить:

- В статьях, где демонстрируется пары "аномалия в b-value → сильное землетрясение" сроки между этими событиями варьируются от 3 месяцев до нескольких лет. Однако, модель не способна это никак учесть или дать оценку сроку.
- Прекурсор основывается на пуассоновском распределении числа землетрясений. Однако, в силу существования форшоков и афтершоков (заметим, что определения этих терминов, основанного на физике процесса – нет) реальный процесс не представляет собой процесс переключения параметра интенсивности между постоянными

значениями. Часто встречающийся подход к решению этой проблемы — разделение каталог на независимые землетрясения и афтершоки. После разделения последние исключаются из рассмотрения. В тоже время, есть работы, показывающие, что такое редактирование каталога создает искусственные аномалии в данных.

2.2 Распределение статистики

При использовании *b-value* все статьи ссылаются на оригинальную статью [7], которая уже не доступна. Так как мы будем использовать эту статистику, приведем вывод ее распределения.

Покажем вспомогательное утверждение. Пусть $f(x; \theta)$ семейство распределений, с носителем x . Рассмотрим $s(x; \theta) = \frac{\partial}{\partial \theta} \log f(x; \theta)$.

$$\begin{aligned} 1 &= \int f(x; \theta) dx \\ 0 &= \frac{\partial}{\partial \theta} \int f(x; \theta) dx = \int \frac{\partial}{\partial \theta} f(x; \theta) dx = \\ &= \int \frac{\partial}{\partial \theta} \log f(x; \theta) f(x; \theta) dx = \mathbb{E}_\theta s(x; \theta) \end{aligned}$$

Тогда для экспоненциального семейства распределений:

$$\begin{aligned} \log f(x; \theta) &= l_\theta = \langle \theta, T(x) \rangle - c(\theta) \\ \frac{\partial}{\partial \theta} l_\theta &= \langle \mathbf{1}, T(x) \rangle - \frac{\partial}{\partial \theta} c(\theta) \\ \mathbb{E} \left(\langle \mathbf{1}, T(x) \rangle - \frac{\partial}{\partial \theta} c(\theta) \right) &= 0 \\ \mathbb{E} T(x)_j &= \frac{\partial}{\partial \theta_j} c(\theta) \end{aligned}$$

Экспоненциальное распределение входит в экспоненциальное семейство распределений. Тогда через дифференцирование константы найдем *MLE* оценку и информацию Фишера. Также воспользуемся асимптотически нормальным распределением *MLE* оценки.

$$f(M, b) = b \exp(-b(M - m_0))$$

$$\hat{b} = \frac{1}{\overline{M} - M_0}$$

$$\sqrt{n}(\hat{b} - b) \sim \mathcal{N}(0, b^2)$$

3 RTL прекурсор

3.1 Обзор

Энергия от деформация при движении плит распространяется на поверхность как волны. Отсюда можно предположить, что функция отражающая "сейсмичность" будет непрерывна относительно координат. Введем такую функцию

$$RTL(x, y, z, t) = R(x, y, z, t) \cdot T(x, y, z, t) \cdot L(x, y, z, t)$$

Каждая из компонент прекурсора улавливает изменения соответственно в пространственной, временной кластеризации землетрясения а также в целом уровень сейсмической активности в окрестности точки (x, y, z, t) .

$$\text{Spatial-Temporal} \left\{ \begin{array}{l} R(x, y, z, t) = \left[\sum_{i \in \mathcal{E}} \exp \left(-\frac{r_i}{r_0} \right) \right], \quad r_i = \|(x, y, z)_{\text{current}} - (x, y, z)_{\text{earthquake } i}\|_2 \\ T(x, y, z, t) = \left[\sum_{i \in \mathcal{E}} \exp \left(-\frac{t - t_i}{t_0} \right) \right] \\ L(x, y, z, t) = \left[\sum_{i \in \mathcal{E}} \exp \left(\frac{l_i}{l_0} \right) \right], \\ \text{for Japan emperical relation } \log l_i = 0.5M_i - 1.8 \end{array} \right.$$

Окрестность \mathcal{E} для землетрясения (x, y, z, t) определяется как все землетрясения, попадающие в пространственно-временной цилиндр:

$$\begin{aligned} M_i &> M_{\min} \\ r_i &< R_{\max} = 2r_0 \\ 0 &< t - t_i < T_{\max} = 2t_0 \end{aligned}$$

В литературе не описывается никаких способов автоматического подбора гиперпараметров, задающих окрестностей. По-сути, задача такая же как выбор оптимальной ширины ядра для восстановления плотности и можно использовать *LOOCV*. Алгоритм оказывается очень чувствительным к гиперпараметрам.

Следуя статье [8], как прекурсоры также рассмотрим усредненное значение RTL

$$Q(x, y, z, t, \Delta t) = \frac{1}{m} \sum_{i=1}^m RTL(x, y, z, t_i), \quad t_i \in [t, \Delta t]$$

Кроме этого, я вычитал среднее и скользящее среднее. Это сделало прекурсор менее шумным.

3.2 Определение критических значений

У данного прекурсора нет вероятностного смысла, соответственно нельзя определить критические квантили. Поэтому возможны несколько стратегий по определению аномальных значений:

- One-class SVM обученный по фиксированному промежутку $[0; T]$
- Эмпирические квантили гистограммы, которую можно перестраивать в скользящем окне
- Критические квантили остатков модели $RTL_t = f(RTL_{t-1}, \dots, (RTL_{t-k}) + \varepsilon_t$, в предположении их независимости и нормальности.

Я использовал последний способ. В качестве модели $f()$ я использовал SVM с `rbf` ядром и линейную регрессию, в обоих случаях использовалась l_1 регуляризация.

4 Новый прекурсор на основе b-value

В прошлом пункте мы рассмотрели прекурсор b-value. Его вероятностная модель – экспоненциальное распределение времени между магнитудами землетрясений. Переходя в такой модели к количеству землетрясений, мы получим распределением Пуассона. Такой подход позволит нам параметризовать λ как линейную функцию от ковариатов x . Изменяя структуру x , мы получаем различные прекурсоры.

- Прекурсор, как аномалии по временной оси $\Leftarrow x = (\lambda_{t-1}, \dots, \lambda_{t-k})$
- Прекурсор, как аномалии в пространственном поле $\Leftarrow x = (\lambda_1, \dots, \lambda_k)$, индекс по k соседям

Кроме этого, мы добавим в модель ядерный подход, что сделает её гибче.

4.1 Вероятностная модель

$$\begin{aligned}\lambda(x) &= x^T \beta \\ \mu(x) &= \exp \lambda(x) \\ y &\sim \text{Poisson}(\mu(x))\end{aligned}$$

Введем априорное распределение $\beta \sim \mathcal{N}(0, \Sigma_\beta)$. Наша цель получить апостериорное распределение

$$p(\beta|X, y) \propto p(y|X, \beta)p(\beta)$$

В силу не сопряженности модели, нормировочная константа нам неизвестна, а интеграл $\int p(y|X, \beta)p(\beta)d\beta$ в аналитическом виде не берется. В такой ситуации, возможны два пути:

- МСМС семплирование
- аппроксимация распределения

Нам необходимо решение в замкнутом виде, т.к. мы хотим последовательно рассчитывать значение прекурсора с приходом новых данных, а также хочется иметь трактуемое критическое значение прекурсора в виде квантиля распределения. Поэтому мы будем использовать аппроксимацию и получим решение в замкнутом виде.

4.2 Вероятностный вывод

Выпишем правдоподобие для наблюдаемых данных $D = (y_i, x_i)_{i=1}^N$ в соответствии с введенной выше моделью.

$$\begin{aligned} p(Y|X, \beta) &= \prod_{i=1}^N \frac{1}{y_i!} \mu(x_i)^{y_i} \exp(-\mu(x_i)) = \\ &= \prod_{i=1}^N \frac{1}{y_i} \underbrace{\frac{1}{(y_i - 1)!} \exp(\lambda(x_i)) \exp(-\exp(\lambda(x_i)))}_{\text{log-Gamma distribution}} \end{aligned}$$

Как показано в работе [9] лог-гамма распределение может быть аппроксимированно в виде:

$$\text{logGamma}(a, b) \approx \mathcal{N}(\log a + \log b, a^{-1})$$

Используя эту аппроксимацию, продолжим вывод:

$$\begin{aligned} \prod_{i=1}^N \frac{1}{y_i} \underbrace{\frac{1}{(y_i - 1)!} \exp(\lambda(x_i)) \exp(-\exp(\lambda(x_i)))}_{\text{log-Gamma distribution}} &= \\ &= \prod_{i=1}^N \frac{1}{y_i} \mathcal{N}(\lambda(x_i) | \log y_i, y_i^{-1}) = \\ &= (2\pi)^{-\frac{N}{2}} |\Sigma_y|^{\frac{1}{2}} \exp \left[-\frac{1}{2} (X\beta - \log(y))^T \Sigma_y^{-1} (X\beta - \log(y)) \right], \text{ где } \Sigma_y = \text{diag} \left(\frac{1}{y_i} \right) \end{aligned}$$

Теперь правдоподобие и априорное распределение для β сопряжены, чем мы и воспользуемся:

$$\begin{aligned} \log p(\beta|X, y) &\propto \log p(y|X, \beta) + \log p(\beta) \propto \\ &- \frac{1}{2} (X\beta - y)^T \Sigma_y^{-1} (X\beta - \log y) - \frac{1}{2} \beta^T \Sigma_\beta^{-1} \beta = Q(\beta) \\ \nabla Q(\beta) &= (X^T \Sigma_y^{-1} X + \Sigma_\beta^{-1}) \beta - X^T \Sigma_y^{-1} X \log y = 0 \\ \beta^* &= \mu'_\beta = (X^T \Sigma_y^{-1} X + \Sigma_\beta^{-1})^{-1} X^T \Sigma_y^{-1} \log y, \text{ где } \Sigma_y = \text{diag} \left(\frac{1}{y_i} \right) \\ \Sigma'_\beta &= \nabla \nabla Q(\beta) = (X^T \Sigma_y^{-1} X + \Sigma_\beta^{-1})^{-1} \end{aligned}$$

Таким образом, мы получили апостериорное распределение:

$$p(\beta|X, y) \sim \mathcal{N}(\mu'_\beta, \Sigma'_\beta)$$

$$\mu'_\beta = (X^T \Sigma_y^{-1} X + \Sigma_\beta^{-1})^{-1} X^T \Sigma_y^{-1} \log y, \text{ где } \Sigma_y = \text{diag} \left(\frac{1}{y_i} \right)$$

$$\Sigma'_\beta = (X^T \Sigma_y^{-1} X + \Sigma_\beta^{-1})^{-1}$$

Тогда распределение $\lambda(x_*)$, где x_* обозначено новое наблюдение:

$$\lambda(x_*) = x_*^T \beta \sim \mathcal{N} \left(x_*^T \mu'_\beta, x_*^T \Sigma'_\beta x_* \right)$$

4.3 Ядровая версия

Получим ядровую версию. Для этого придется ограничиться $\Sigma'_\beta = I$.

$$\mu_\lambda = x_*^T (X^T \Sigma_y^{-1} X + \Sigma_\beta^{-1})^{-1} X^T \Sigma_y^{-1} \log y$$

Применим тождество Шермана-Моррисона-Вудбери

$$(X^T \Sigma_y^{-1} X + I)^{-1} X^T \Sigma_y^{-1} = X^T (X X^T + \Sigma_y)^{-1}$$

$$\mu_\lambda = x_*^T X^T (X X^T + \Sigma_y)^{-1} \log y$$

Тогда естественным образом, мы можем заменить матрицу Грамма на ядровую матрицу. Введем обозначения.

$$k(\cdot, \cdot) = \langle \phi(\cdot), \phi(\cdot) \rangle, \text{ ядровая функция}$$

$$k_* = [k(x_*, x_1) \dots, k(x_*, x_N)]^T$$

$$K = (K_{ij})_{i,j=1,\dots,N} = (k(x_i, x_j))_{i,j=1,\dots,N}$$

Тогда

$$\lambda(x_*) \sim \mathcal{N}(\mu_\lambda^k, \Sigma_\lambda^k)$$

$$\mu_\lambda^k = k_*^T (K + \Sigma_y)^{-1} \log y$$

$$\Sigma_\lambda^k = k(x_*, x_*) - k_*^T (K + \Sigma_y)^{-1} k_*$$

Таким образом, мы получили ядровую версию.

4.4 Байесовский прогноз

Также как прекурсор можно использовать временной ряд невязки между реальным числом землетрясений и нашим предсказанием. Заметим, что в целом, на не важна точность прогноза сама по себе, главное чтобы она была достаточной для того, чтобы уловить разладку перед землетрясениями.

$$p(y_*|x_*, X, y) = \int \underbrace{p(y_*|\lambda_*)}_{\text{Poisson}} \underbrace{p(\lambda_*|x_*, X, y)}_{\text{Normal}}$$

теперь используем аппроксимацию в обратную сторону, чтобы получить сопряженную пару Gamma-Poisson

$$\begin{aligned} \mu_* = \exp(\lambda_*) &\Rightarrow \lambda_* = \log \mu^*, \quad \lambda_* \sim \text{Normal} \approx \text{LogGamma} \\ &\Rightarrow \mu^* \sim \text{Gamma}(a_*, b_*) \\ a_* &= \sigma_\lambda^{-2} \\ b_* &= \sigma_\lambda^2 \exp(\mu_\lambda) \end{aligned}$$

Сопряженная пара Gamma-Poisson дает Negative-Binomial распределение:

$$y_*|x_*, X, y \sim \text{NegBin}(\exp(\mu_\lambda), \sigma_\lambda^2)$$

5 Траектории землетрясений как смесь регрессий

5.1 Теория вариационного вывода

Далее для смеси регрессий будет использоваться вариационный вывод [10]. Кратко сформулируем и выпишем алгоритм вариационного вывода в общем виде

Пусть модель:

$$p(X, Z, \theta) = p(X|Z)p(Z|\theta)p(\theta)$$

Цель:

$$p(Z, \theta|X) = \frac{p(X, Z, \theta)}{\int p(X, Z, \theta) dZ d\theta}$$

Самое сложное это получить $\int p(X, Z, \theta) dZ d\theta = p(X)$. Мы не можем вычислить $\int p(X, Z, \theta) dZ d\theta$ напрямую. Тогда мы ограничиваем множество функций, в котором ищем решение как $q(Z, \theta) = q_z(Z)q_\theta(\theta)$. Такое ограничение называется "mean field approximation"

$$\mathcal{F}(q) = \mathbb{E}_q \log p(X, Z|\theta) + \mathbb{E}_q \log p(\theta) - \mathbb{E}_q \log q(Z, \theta) = \mathbb{E}_q \log p(X, Z|\theta) + \mathbb{E}_{q_\theta} \log p(\theta) - \mathbb{E}_{q_\theta} \log q_\theta(\theta)$$

Условия первого порядка:

$$\frac{\delta}{\delta q_\theta} \mathcal{L} = \mathbb{E}_{q_z} \log p(X, Z|\theta) + \log p(\theta) - (\log q_\theta(\theta) + 1) + \lambda_\theta = 0$$

$$\frac{\delta}{\delta q_z} \mathcal{L} = \mathbb{E}_{q_\theta} \log p(X, Z|\theta) - (\log q_z(Z) + 1) + \lambda_z = 0$$

$$\frac{\partial}{\partial \lambda_\theta} \mathcal{L} = \int q_\theta(\theta) d\theta - 1 = 0$$

$$\frac{\partial}{\partial \lambda_z} \mathcal{L} = \int q_z(Z) dZ - 1 = 0$$

Тогда для q_θ :

$$q_\theta(\theta) = p(\theta) \exp(\mathbb{E}_{q_z} \log p(X, Z|\theta) - 1 + \lambda_\theta)$$

$$\lambda_\theta = 1 - \log \int p(\theta) \mathbb{E} \exp(\mathbb{E}_{q_z} \log p(X, Z|\theta)) d\theta$$

Окончательно

$$q_\theta(\theta) = \frac{p(\theta) \mathbb{E} \exp(\mathbb{E}_{q_z} \log p(X, Z|\theta))}{\int p(\theta) \mathbb{E} \exp(\mathbb{E}_{q_z} \log p(X, Z|\theta)) d\theta}$$

В силу симметрии:

$$q_z(Z) = \frac{\mathbb{E} \exp(\mathbb{E}_{q_\theta} \log p(X, Z|\theta))}{\int \mathbb{E} \exp(\mathbb{E}_{q_\theta} \log p(X, Z|\theta)) dZ}$$

5.2 Смесь регрессий

- $D = \{(y_i, x_i)\}_{i=1}^N$
- $y_i \in \mathbb{R}, x_i \in \mathbb{R}^5$
- Генеративная модель для y : $P(y_i|\theta) = \sum_{k=1}^K \pi_k \mathcal{N}(y_i|x_i^T \beta_k, \tau_k)$ Т.е. мы моделируем K независимых траекторий, каждая из которых описывается как $y_i = x_i^T \beta_k + \varepsilon_i^k$. Однако, мы не наблюдаем латентные переменные, поэтому y описывается как смесь.
- Введем латентные переменные принадлежности к кластеру Z где для z_n используется стандартная K-coding схема: $\{0; 1\}^K$ and $\sum_k z_n k = 1$
- Для каждого β_k введем нормальное априорное распределение $\mathcal{N}(\beta_k|\mu_\beta^0, \Sigma_0)$ и для τ_k Inverse-Gamma $\text{IG}(\tau_k|a_0, b_0)$
- Мы хотим, чтобы число траекторий-кластеров определялось автоматически. Поэтому мы используем в качестве априорного распределения для весов разреженный Dirichlet distribution. Начиная с заведомого большего числа кластеров, итеративно мы сойдемся к $k, k < K$.

Итак, модель:

$$\prod_{n=1}^N \prod_{k=1}^K [\mathcal{N}(y_n|x_n^T \beta_k, \tau_k)]^{z_{nk}} \prod_{n=1}^N \prod_{k=1}^K [\pi_k]^{z_{nk}} \text{Dir}(\pi|\alpha) \prod_{k=1}^K \mathcal{N}(\beta_k|\mu_\beta^0, \Sigma_0) \text{IG}(\tau_k|a_0, b_0)$$

5.3 Отличие от других подходов

Рассмотрим несколько статей, решающих такую же задачу разделения траекторий и укажем на отличия нашего решения. В качестве вывода мы использовали вариационный подход, получив решение в аналитическом виде. Также мы использовали априорные распределения на параметры модели в качестве регуляризации и для автоматического определения числа кластеров. В статьях же в основном используется MCMC семплинг, а для выбора числа кластеров информационный критерий.

- Dirichlet Prior + MCMC : "Estimating Mixtures of Regressions" Merrilee HURN , Ana JUSTEL , and Christian P. ROBERT [11]
- Dirichlet Process + MCMC: "Dirichlet Process Mixtures of Generalized Linear Models" Lauren A. Hannah, David M. Blei, Warren B. Powell [12]
- BIC критерий + EM алгоритм: "Extending the Akaike Information Criterion to Mixture Regression Models" Prasad A. NAIK, Peide SHI, and Chih-Ling T [13]

5.4 Вероятностный вывод

Будем строить вариационное приближение в классе плотностей, факторизующихся как:

$$p(Z, \pi, \beta, \tau | D) = q(Z|X)q(\pi|X)q(\beta)q(\tau)$$

Также, так как мы считаем, что выборка *iid*, то все распределения факторизуются по наблюдениям.

$$q(\beta_k)$$

$$\begin{aligned} \log q(\beta_k) &\propto \mathbb{E}_{q(\beta_k)} \left[\sum_{n=1}^N z_{nk} \log \mathcal{N}(y_n | x_n^T \beta_k, \tau_k) + \log \mathcal{N}(\beta_k | \mu_\beta^0, \Sigma_0) \right] = \\ &= \mathbb{E}_{q(\beta_k)} \left[\sum_{n=1}^N z_{nk} \left(-\frac{1}{2} \frac{1}{\tau_k} (y_n - x_n^T \beta_k)^T (y_n - x_n^T \beta_k) \right) - \frac{1}{2} (\beta_k - \mu_\beta^0)^T \Sigma_0^{-1} (\beta_k - \mu_\beta^0) \right] = \\ &= \mathbb{E}_{q(\beta_k)} \left[(Y - X \beta_k)^T \text{diag} \left(\frac{z_{nk}}{\tau_k} \right) (Y - X \beta_k) - \frac{1}{2} (\beta_k - \mu_\beta^0)^T \Sigma_0^{-1} (\beta_k - \mu_\beta^0) \right] = \\ &= -\frac{1}{2} (Y - X \beta_k)^T W_k (Y - X \beta_k) - \frac{1}{2} (\beta_k - \mu_\beta^0)^T \Sigma_0^{-1} (\beta_k - \mu_\beta^0), \\ \text{where } W_k &= \text{diag} \left(\frac{\mathbb{E}_{q(z)} z_{nk}}{\mathbb{E}_{q(\tau)} \tau_k} \right) \end{aligned}$$

Поэтому, $q(\beta_k) \sim \mathcal{N}(\beta_k | \mu_\beta^k, \Sigma_k)$. Чтобы найти параметры, мы максимизируем квадратичную форму, также используем $q(\tau, z) = q(z)q(\tau)$

$$\begin{aligned}
Q(\beta_k) &= -\frac{1}{2}(Y - X\beta_k)^T W_k (Y - X\beta_k) - \frac{1}{2}(\beta_k - \mu_\beta^0)^T \Sigma_0^{-1} (\beta_k - \mu_\beta^0) \\
\nabla Q(\beta_k) &= - (X^T W_k (X\beta_k - Y) + \Sigma_0^{-1} (\beta_k - \mu_\beta^0)) = 0 \\
\beta_k^* &= \mu_\beta^k = (XW_k X + \Sigma_0^{-1})^{-1} [X^T Y + \Sigma_0^{-1} \mu_\beta^0] \\
W_k &= (XW_k X + \Sigma_0^{-1})^{-1} \\
\text{where } W_k &= \text{diag} \left(\frac{\mathbb{E}_{q(z)} z_{nk}}{\mathbb{E}_{q(\tau)} \tau_k} \right)
\end{aligned}$$

$$q(\tau_k)$$

$$\begin{aligned}
\log q(\tau_k) &\propto \mathbb{E}_{q(\tau_k)} \left[\sum_{n=1}^N z_{nk} \left(-\frac{n}{2} \log \tau_k - \frac{1}{2} \frac{1}{\tau_k} (y_n - x_n^T \beta_k)^T (y_n - x_n^T \beta_k) \right) - (a_0 + 1) \log \tau_k - \frac{b_0}{\tau_k} \right] = \\
&= - \left(\frac{n}{2} \sum_{n=1}^N \mathbb{E}_{q(z)} z_{nk} + a_0 + 1 \right) \log \tau_k - \frac{1}{2} \frac{1}{\tau_k} \left(b_0 + \sum_{n=1}^N \mathbb{E}_{q(z, \beta)} z_{nk} (y_n - x_n^T \beta_k)^T (y_n - x_n^T \beta_k) \right)
\end{aligned}$$

Отсюда $q(\tau_k) \sim \text{IG}(\tau_k | a_k, b_k)$. Используя $q(z, \beta) = q(z)q(\beta)$:

$$a_k = a_0 + \frac{n}{2} \sum_{n=1}^N \mathbb{E}_{q(z)} z_{nk}$$

$$b_k = b_0 + \sum_{n=1}^N \mathbb{E}_{q(z, \beta)} z_{nk} (y_n - x_n^T \beta_k)^T (y_n - x_n^T \beta_k)$$

$$\sum_{n=1}^N \mathbb{E}_{q(z, \beta)} z_{nk} (y_n - x_n^T \beta_k)^T (y_n - x_n^T \beta_k) = \text{Tr} (Z_k \mathbb{E}_{q(\beta)} (Y - X\beta)(Y - X\beta)^T), \text{ where } Z_k = \text{diag} (\mathbb{E}_{q(z)} z_{nk})$$

, finally:

$$b_k = b_0 + \frac{1}{2} \left[(Y - X\mu_{q(\beta_k)})^T Z (Y - X\mu_{q(\beta_k)}) \right] + \frac{1}{2} \text{Tr} (X^T Z X \Sigma_{q(\beta_k)})$$

$$q(Z)$$

$$\log q(z) \propto \mathbb{E}_{\setminus z} \sum_{n=1}^N \sum_{k=1}^K z_{nk} [\log \mathcal{N}(y_n | x_n^T \beta_k, \tau_k) + \log \pi_k]$$

$$\mathbb{E} z_{nk} = \rho_{nk}$$

$$\begin{aligned} \log \rho_{nk} &= \mathbb{E}_{\setminus z} \sum_{n=1}^N \sum_{k=1}^K z_{nk} [\log \mathcal{N}(y_n | x_n^T \beta_k, \tau_k) + \log \pi_k] = \\ &= \mathbb{E}_{q(\pi)} \log \pi_k - \frac{1}{2} \log 2\pi - \frac{1}{2} \mathbb{E}_{q(\tau_k)} \log \tau_k - \mathbb{E}_{q(\tau_k)} \tau_k^{-1} ((y_n - x_n^T \mu_{q(\beta_k)} x_n^T)^2 + x_n^T \Sigma_{q(\beta_k)} x_n) \end{aligned}$$

$$q(\pi)$$

$$\log q(\pi) \propto \mathbb{E}_{\setminus \pi} \sum_{n=1}^N \sum_{k=1}^K z_{nk} \log \pi_k + \log \text{Dir}$$

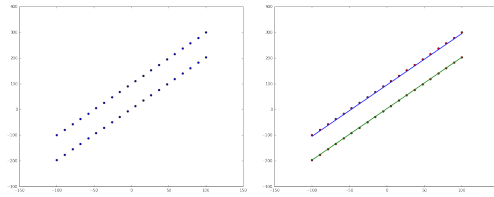
$$\text{Так что, } q(\pi) \sim \text{Dir}(\pi | \alpha_0 + \sum_{n=1}^N \mathbb{E}_{q(z)} z_{nk})$$

Все необходимые матожидания это матожидания достаточных статистик распределений из экспоненциального класса и легко выписываются.

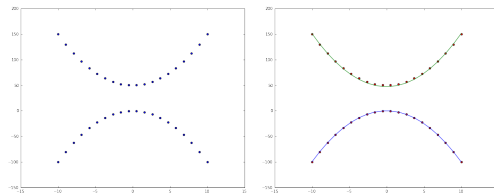
5.5 Иллюстративные примеры на искусственных данных

Так как вариационный алгоритм сходится к локальному максимуму, мы делаем несколько запусков и выбираем тот, который получил наибольшую ELBO.

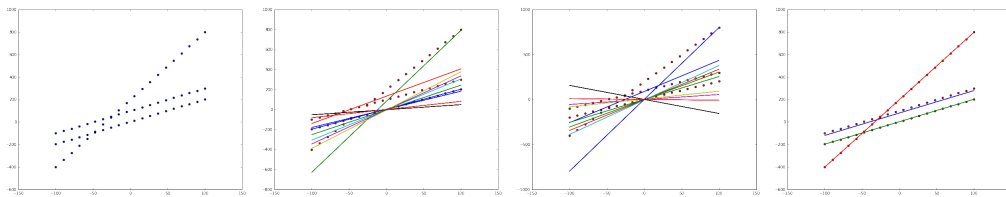
Примеры, показывающие разделение траекторий:



Second example, ground truth and model work MAP estimation of regression:



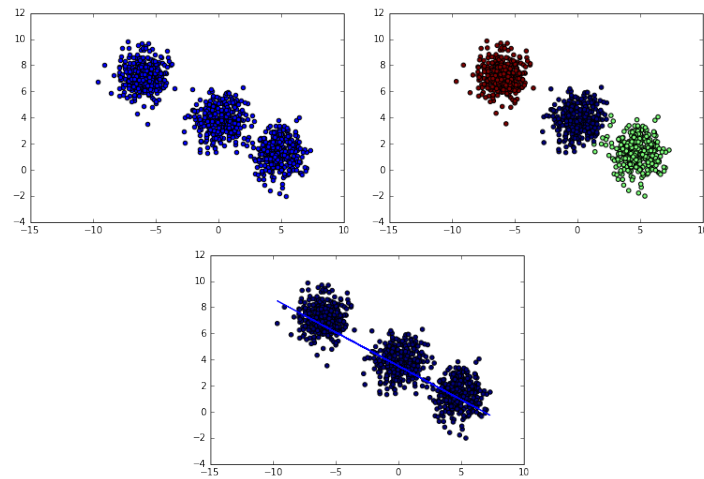
Пример, показывающий автоматический подбор числа кластеров: Стартуем с $K = 24$ кластеров, и Dirichlet prior $\alpha_0^k = 0.1, \forall k$. Сходимся к 3 кластерам, что соответствует механизму генерации данных.



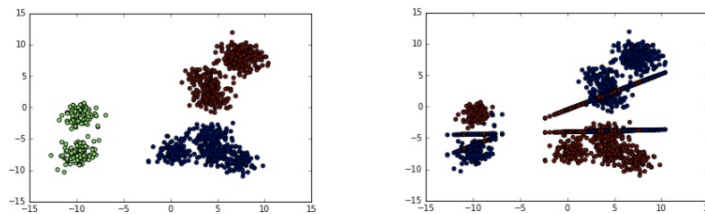
5.6 Кластеризация вдоль траектории

Также мы можем кластеризовать данные вдоль траекторий. Рассмотрим модельные примеры. В каждом из них обычная смесь гауссиан дает избыточное число кластеров по сравнению с кластеризацией вдоль траекторий.

Один пример:



Другой пример:



5.7 Использование для предсказания землетрясений

- Декластеризация каталогов землетрясений Как было отмечено ранее (в секции про b-value прекурсор) многие рассматриваемые в литературе прекурсоры землетрясений опираются на предпосылку о пуассоновском распределении землетрясений в каталоге. В тоже

время это противоречит природе землетрясений: они кластеризуются в результате фошковых и афтершоков. Чтобы исправить ситуацию, нужно выделить такие последовательности и удалить все наблюдения, кроме одного, с наибольшей магнитудой. Как метод выделения таких последовательностей можно использовать предложенную процедуру кластеризации вдоль направлений.

- Выявление типичных траекторий и расчет статистик внутри них. Однако мы поступим иначе и не будем удалять наблюдения. Вместо этого рассчитаем такие направления по месячным подвыборкам и будем рассчитывать прекурсоры в таких кластерах.

6 Предсказание землетрясения

Мы рассмотрели пару самых популярных прекурсоров и предложили свой. Кроме этого, для прекурсора *RTL* задавая различные гиперпараметры можно получить серию прекурсоров. Также был получен метод кластеризации землетрясений.

Опишем наш подход к предсказанию землетрясений:

1. Выделение подсистем: пространственных кластеров
2. Расчет прекурсоров в каждом кластере, регрессия прекурсоров $\setminus k$ кластеров на значения прекурсора в кластере k
3. Выделение наилучшего подмножества прекурсоров и решающего правила над ними

Нам осталось описать алгоритм, который будет выделять наилучшего подмножества прекурсоров.

6.1 Агрегация различных прекурсоров

Значения различных прекурсоров представляют собой бинарные ответы, есть аномалия, по-мнению прекурсора, или нет. Ответы можно агрегировать как логическое "и" логическое "или" либо как голосование. Опишем алгоритм выбора наилучшего подмножества:

Пусть $(X_{ft}) \in \{0; 1\}$ бинарные значения прекурсора f из множества всех прекурсоров $f \in F$ и $(y_t)_{t=1}^T \in \{0; 1\}$ разметка произошло ли сильно землетрясение.

- Разобьем временную ось на K не пересекающихся интервалов
- Обозначим K^+ число успешных тревог, и за K^- ложных
- Для каждого $f \in F$ подсчитаем:
 - Число успешных тревог в общем и уникальных S_f^+, U_f^+ соответственно
 - Число ложных тревог в общем и уникальных S_f^-, U_f^- соответственно

- Рассчитаем вероятности успешного и ложного срабатываний

$$p^* = \frac{S^*}{K^*}, * \in \{+, -\}$$
- Рассчитаем P-value $p_f^+ = p_f^-$
- Если гипотеза о равенстве отвергается, то такой предиктор f остается.

7 Результаты

7.1 Япония

Данные

Землетрясения в окрестности Японии с 1990 по 2015 год. Целевыми землетрясениями считали землетрясения с магнитудой более 6. Обучающая выборка содержала 500 целевых событий и 185210 землетрясений малой магнитуды.

Отобранные прекурсоры

Перебор ввелся по множествам мощности 3, решающее правило: логическое или

- RTL, с параметрами окна $R_c = 200, T_c = 365, M_c = 3$
- Предложенный прекурсор, в качестве ковариатов использовались среднее магнитуды у 5 ближайших соседей
- Невязка регрессии RTL ближайщего кластера против RTL в остальных кластерах

Ядерные варианты предложенного прекурсора, к сожалению не вошли.

δ_c временной между предективным окном и целевым событием. Длина предиктивного окна: 300 дней.

Результат

Таблица 1: Результат

δ_c	всего событий	всего срабатываний	предсказано успешно	ложные срабатывания
15	80	1765	64	1701
20	80	1261	71	1190
30	80	985	70	915

7.2 Италия

Данные

Землетрясения в окрестности Италии с 2000 по 2014 год. Целевыми землетрясениями считали землетрясения с магнитудой более 4.5. Обучающая выборка содержала 100 целевых событий и 2673 землетрясений малой магнитуды.

Отобранные прекурсоры

Перебор ввелся по множествам мощности 3, решающее правило: логическое или

- RTL, с параметрами окна $R_c = 150, T_c = 365, M_c = 1.$, т.е. фактически отсечки по минимальной магнитуде не было
- RTL, с параметрами окна $R_c = 150, T_c = 365, M_c = 3.$
- Предложенный прекурсор, в качестве ковариатов использовались средние магнитуды у 3 ближайших соседей

δ_c временной между предективным окном и целевым событием. Длина предиктивного окна: 300 дней.

Результат

Таблица 2: Результат				
δ_c	всего событий	всего срабатываний	предсказано успешно	ложные срабатывания
15	20	49	13	36
20	20	53	12	41
30	20	46	18	28

7.3 Выводы

Не строго можно утверждать, что прекурсоры обладают предсказательной силой. Однако, на примере Японии, хорошо видно, что нужны более частотные данные, чем каталог землетрясений. В Италии, где места землетрясений и их количество гораздо разреженнее,

относительно Японии, картина предсказаний значительно лучше. RTL очень мощный прекурсор. Надо адаптировать его как ядро и попробовать использовать в ядровых методах

Список литературы

- [1] Working Group on California Earthquake Probabilities. Earthquakes probabilities in the san francisco bay region: 2002 to 2031. *U.S. Geological Survey Open File Report*, 2003.
- [2] B. GUTENBERG and C. RICHTER. Seismicity of the earth and associated phenomena. *Princeton University Press, Princeton, New Jersey, 2nd edition*, page 310, 1954.
- [3] Yuan Gao Stuart Crampin. The physics underlying gutenbergrichter in the earth and in the moon. *Journal of Earth Science*, 26(1):134–139, 2015.
- [4] Allen C. P. Amand C. Richter and J. Nordquist. Relation between seismicity and geological structure in the southern california region. *Bull. Seismol. Soc. Am*, 1965.
- [5] Mori and R.E. Abercrombie. Depth dependence of earthquake frequency-magnitude distributions in california: Implication for the rupture initiation. *Geophys Res*, 1997.
- [6] R.P. Young Urbancic T. I. C-I. Trifu J. M. Long. Space-time correlation of b values with stress release. *PAGEOPH*, pages 449–462, 1992.
- [7] Aki K. Maximum likelihood estimate of b in the formula $\log n = a - bm$ and its confidence limits. *Bulletin of the Earthquake Research Institute*, 43:237–239, 1965.
- [8] Qinghua Huang. Seismicity pattern changes prior to large earthquakes-an approach of the rtl algorithm. *TAO*, 15(3):469–491, 2004.
- [9] Kendall D.G Bartlett M.S. The statistical analysis of variance-heterogeneity and logarithmic transformation. *Jn.R. Statist.Soc. Suppl.*, 8, 128, 1946.
- [10] David M. Blei Alp Kucukelbir Jon D. McAuliffe. Variational inference: A review for statisticians. *arXiv*, 2016.
- [11] Merrilee HURN Ana JUSTEL Christian P. ROBERT. Estimating mixtures of regressions. *Am Statist Assoc*, 2008.

- [12] Lauren A. Hannah David M. Blei Warren B. Powell. Dirichlet process mixtures of generalized linear models. *Journal of Machine Learning Research* 1, 2011.
- [13] Prasad A NAIK Peide SHI and Chih-Ling T. Extending the akaike information criterion to mixture regression models. *Am Statist Assoc*, 2007.