MM-алгоритм и медианная регрессия

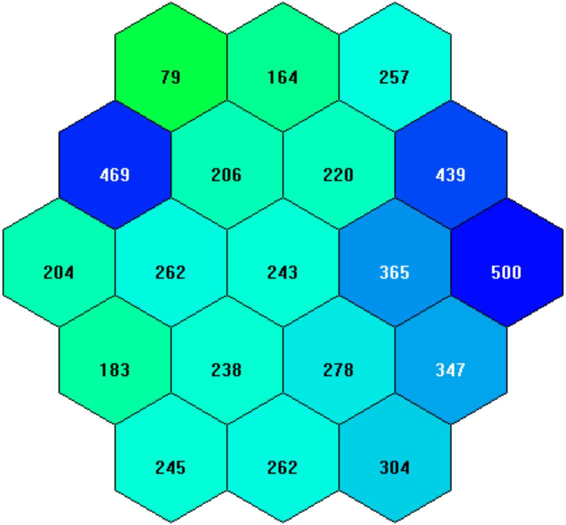
Ветров Д. П., Кропотов Д. А., Касперский И. Е. [1]

[1] МГУ, ВМиК, каф. ММП

23 ноября 2011

Тактильный механорецептор (пальпатор)

Тактильный механорецептор (пальпатор): расположение каналов



Тактильный механорецептор (пальпатор)

Как проверять работоспособность приборов?

Тактильный механорецептор (пальпатор)

Как проверять работоспособность приборов?

- Взять специально обученного человека

Тактильный механорецептор (пальпатор)

Как проверять работоспособность приборов?

- Взять специально обученного человека
  Плюсы:
  - Не нужно ничего делать

  Минусы:
  - Человек не может всегда создавать одинаковую силу нажатия
  - Нет гарантии вертикального нажатия
  - Трудно выявить взаимодействие каналов прибора друг с другом
  - Не требуется машинное обучение

Тактильный механорецептор (пальпатор)

Как проверять работоспособность приборов?

- Автоматически, создав тестирующую установку

Тактильный механорецептор (пальпатор)

Как проверять работоспособность приборов?

- Автоматически, создав тестирующую установку
  Плюсы:
  - Техника обеспечит воспроизводимость и точность результатов
  - Можем создать сложную схему для выявления любого рода неисправностей

  Минусы:
  - Шум и выбросы во входных данных

«Тактильный дисплей»

«Тактильный дисплей»

- Позволяет оценить текущую картину на пальпаторе «вживую»
- Подходит только для грубых качественных оценок

Тестирующая установка «Кофеварка»

Тестирующая установка «Кофеварка»

- Позволяет осуществить вертикальное нажатие заданной силы

Тестирующая установка «Кофеварка»

- Позволяет осуществить вертикальное нажатие заданной силы
- Позволяет найти зависимость показаний пальпатора от реального давления

Тестирующая установка «Кофеварка»

- Позволяет осуществить вертикальное нажатие заданной силы
- Позволяет найти зависимость показаний пальпатора от реального давления
- Даёт только общую информацию обо всей совокупности каналов

Тестирующая установка «Кофеварка»

- Позволяет осуществить вертикальное нажатие заданной силы
- Позволяет найти зависимость показаний пальпатора от реального давления
- Даёт только общую информацию обо всей совокупности каналов
- Вопрос синхронизации с пальпатором

Тестирующая установка «Реактор»

Тестирующая установка «Реактор»

- Осуществляет подачу давления на группы каналов

Тестирующая установка «Реактор»

- Осуществляет подачу давления на группы каналов
- Поэтому позволяет обнаруживать межканальные взаимодействия

Тестирующая установка «Реактор»

- Осуществляет подачу давления на группы каналов
- Поэтому позволяет обнаруживать межканальные взаимодействия
- Подходит для исследования рабочих каналов

Тестирующая установка «Реактор»

- Осуществляет подачу давления на группы каналов
- Поэтому позволяет обнаруживать межканальные взаимодействия
- Подходит для исследования рабочих каналов
- Тестируется только датчик пальпатора

Процедура тестирования

- Находим рабочие каналы (на «Реакторе»)

Процедура тестирования

- Находим рабочие каналы (на «Реакторе»)
- Их показания переводим в единую шкалу (на «Реакторе»)

Процедура тестирования

- Находим рабочие каналы (на «Реакторе»)
- Их показания переводим в единую шкалу (на «Реакторе»)
- Находим зависимость между реальным давлением и показаниями пальпатора (на «Кофеварке»)

ММ-алгоритм: шаг первый

Суть: мажорирование - минимизация.

MM-алгоритм: шаг первый

Суть: мажорирование - минимизация.

Хотим минимизировать $L(\theta) : \mathbb{R}^p \to \mathbb{R}$.

MM-алгоритм: шаг первый

Суть: мажорирование - минимизация.

Хотим минимизировать $L(\theta) : \mathbb{R}^p \to \mathbb{R}$.

Шаг первый: для текущей итерации алгоритма θ^k создаём функцию $Q(\theta|\theta^k) : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}$, такую, что выполнены два условия:

$$Q(\theta^k|\theta^k) = L(\theta^k) \tag{1}$$

$$Q(\theta|\theta^k) \geq L(\theta^k), \forall\theta \tag{2}$$

Эта функция мажорирует $L(\theta)$.

MM-алгоритм: шаг второй

Суть: мажорирование - минимизация.

Хотим минимизировать $L(\theta) : \mathbb{R}^p \to \mathbb{R}$.

Шаг второй: минимизировать функцию $Q(\theta|\theta^k)$ по θ. Надо выбирать удобную Q.

MM-алгоритм: шаг второй

Суть: мажорирование - минимизация.

Хотим минимизировать $L(\theta) : \mathbb{R}^p \to \mathbb{R}$.

Шаг второй: минимизировать функцию $Q(\theta|\theta^k)$ по θ. Надо выбирать удобную Q.

За следующую итерацию примем θ^{k+1}, минимизирующее $Q(\theta|\theta^k)$.

MM-алгоритм: шаг второй

Суть: мажорирование - минимизация.

Хотим минимизировать $L(\theta) : \mathbb{R}^p \to \mathbb{R}$.

Шаг второй: минимизировать функцию $Q(\theta|\theta^k)$ по θ. Надо выбирать удобную Q.

За следующую итерацию примем θ^{k+1}, минимизирующее $Q(\theta|\theta^k)$.

Тогда имеем:

$$L(\theta^{k+1}) \leq Q(\theta^{k+1}|\theta^k) \leq Q(\theta^k|\theta^k) = L(\theta^k)$$

Медианная регрессия

Рассматриваемая задача: регрессия x на y с функцией потерь

$$L(\theta) = \sum_{i=1}^{n} |y_i - x\theta_1 - \theta_2|$$

Медианная регрессия

Рассматриваемая задача: регрессия x на y с функцией потерь

$$L(\theta) = \sum_{i=1}^{n} |y_i - x\theta_1 - \theta_2|$$

Для простоты обозначим $x\theta_1 - \theta_2$ за $f_i(\theta)$, а i-ю невязку за $r_i = r_i(\theta) = y_i - f_i(\theta)$.

Медианная регрессия

Рассматриваемая задача: регрессия x на y с функцией потерь

$$L(\theta) = \sum_{i=1}^{n} |y_i - x\theta_1 - \theta_2|$$

Для простоты обозначим $x\theta_1 - \theta_2$ за $f_i(\theta)$, а i-ю невязку за $r_i = r_i(\theta) = y_i - f_i(\theta)$.

Необходимое условие существования решения:

$$\lim_{\|\theta\| \to \inf} \sum_{i=1}^{n} f_i^2(\theta) = \inf$$

Медианная регрессия

Функции f_i не дифференцируемы в нуле и L может достигать минимума на отрезке.

Медианная регрессия

Функции f_i не дифференцируемы в нуле и L может достигать минимума на отрезке.
Поэтому введём возмущённую функцию

$$\rho_\varepsilon(r) = |r| - \varepsilon \ln(\varepsilon + |r|) \tag{3}$$

Медианная регрессия

Функции f_i не дифференцируемы в нуле и L может достигать минимума на отрезке.
Поэтому введём возмущённую функцию

$$\rho_\varepsilon(r) = |r| - \varepsilon \ln(\varepsilon + |r|) \tag{3}$$

Тогда сумма

$$L_\varepsilon(\theta) = \sum_{i=1}^{n} \rho_\varepsilon(r_i) \tag{4}$$

приближает исходную функцию потерь, причём

$$L_\varepsilon(\theta_\varepsilon^*) - L(\theta^*) \leq -2\varepsilon n \ln(\varepsilon)$$

где θ_ε^* и θ^* — точки минимума функций $L_\varepsilon(\theta)$ и $L(\theta)$ соответственно.

Медианная регрессия

Теперь можно применить ММ-алгоритм к функции $L_\varepsilon(\theta)$.

Медианная регрессия

Теперь можно применить ММ-алгоритм к функции $L_\varepsilon(\theta)$.

Функция, мажорирующая ρ_ϵ для текущей невязки $r^k = r(\theta^k)$:

$$\xi_\varepsilon(r|r^k) = \frac{1}{2}\left(\frac{rr}{\varepsilon + |r^k|} + c\right) \tag{5}$$

где c — константа, выбираемая таким образом, что $\xi_\varepsilon(r^k|r^k) = \rho_\varepsilon(r^k)$.

Медианная регрессия

Теперь можно применить ММ-алгоритм к функции $L_\varepsilon(\theta)$.

Функция, мажорирующая ρ_ϵ для текущей невязки $r^k = r(\theta^k)$:

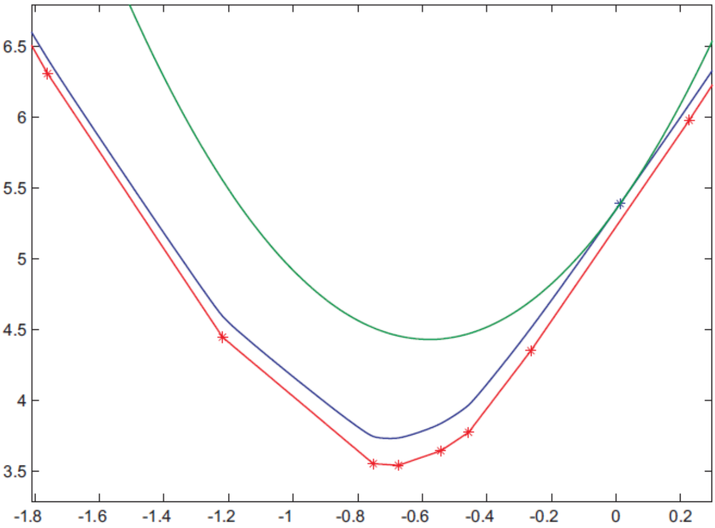
$$\xi_\varepsilon(r|r^k) = \frac{1}{2}\left(\frac{rr}{\varepsilon + |r^k|} + c\right) \tag{5}$$

где c — константа, выбираемая таким образом, что $\xi_\varepsilon(r^k|r^k) = \rho_\varepsilon(r^k)$.

Тогда можно в качестве мажоранты взять

$$Q_\varepsilon(\theta|\theta^k) = \sum_{i=1}^{n} \xi_\varepsilon(r_i, r_i^k)$$

Медианная регрессия



Медианная регрессия

Функция Q_ε квадратична по θ в силу квадратичности по r_i!

Медианная регрессия

Функция Q_ε квадратична по θ в силу квадратичности по r_i!

Для оптимизации этой функции воспользуемся подходом Гаусса-Ньютона.

Медианная регрессия

Функция Q_ε квадратична по θ в силу квадратичности по r_i!

Для оптимизации этой функции воспользуемся подходом Гаусса-Ньютона.

Обозначим для удобства

$$\alpha(r_i) = \frac{r_i}{\sqrt{2(\varepsilon + |r_k|)}} = \beta(x_i) = \frac{y_i - \langle \theta, x_i \rangle}{\sqrt{2(\varepsilon + |r_k|)}}$$

Медианная регрессия

Функция Q_ε квадратична по θ в силу квадратичности по r_i!

Для оптимизации этой функции воспользуемся подходом Гаусса-Ньютона.

Обозначим для удобства

$$\alpha(r_i) = \frac{r_i}{\sqrt{2(\varepsilon + |r_k|)}} = \beta(x_i) = \frac{y_i - \langle \theta, x_i \rangle}{\sqrt{2(\varepsilon + |r_k|)}}$$

Тогда имеем:

$$Q_\varepsilon(\theta|\theta^k) = \sum_{i=1}^{n} \beta^2(x_i) + const = P(\theta) + const$$

Медианная регрессия: метод Гаусса-Ньютона

Пусть $\theta^{k+1} = \theta^k + \Delta$, тогда

$$P(\theta^{k+1}) = P(\theta^k) + \left\| \frac{\partial P}{\partial \theta_i} \right\| \Delta + \frac{1}{2} \Delta^T \left\| \frac{\partial^2 P}{\partial \theta_i \partial \theta_j} \right\| \Delta$$

Медианная регрессия: метод Гаусса-Ньютона

Пусть $\theta^{k+1} = \theta^k + \Delta$, тогда

$$P(\theta^{k+1}) = P(\theta^k) + \left\| \frac{\partial P}{\partial \theta_i} \right\| \Delta + \frac{1}{2} \Delta^T \left\| \frac{\partial^2 P}{\partial \theta_i \partial \theta_j} \right\| \Delta$$

Обозначим матрицу Якоби за

$$J = \left\| \frac{\partial \beta(x_i)}{\partial \theta_j} \right\| = diag\left(- \frac{x_i}{\sqrt{2(\varepsilon + |r_i^k|)}} \right) = -SX,$$

где $X = (x_1, ..., x_n), S = diag(s_1, ..., s_n), s_i = 1/\sqrt{2 * (\varepsilon + |r_i|)}$.

Медианная регрессия: метод Гаусса-Ньютона

Пусть $\theta^{k+1} = \theta^k + \Delta$, тогда

$$P(\theta^{k+1}) = P(\theta^k) + \left\| \frac{\partial P}{\partial \theta_i} \right\| \Delta + \frac{1}{2} \Delta^T \left\| \frac{\partial^2 P}{\partial \theta_i \partial \theta_j} \right\| \Delta$$

Обозначим матрицу Якоби за

$$J = \left\| \frac{\partial \beta(x_i)}{\partial \theta_j} \right\| = diag\left(- \frac{x_i}{\sqrt{2(\varepsilon + |r_i^k|)}} \right) = -SX,$$

где $X = (x_1, ..., x_n), S = diag(s_1, ..., s_n), s_i = 1/\sqrt{2 * (\varepsilon + |r_i|)}$.

Тогда матрицу первых производных можно заменить на $-SX(\beta_1, ...\beta_n) = -SX\vec{\beta}$, а гессиан — на $(-SX)^T(-SX) = X^T S^T SX = X^T S^2 X$.

Медианная регрессия: метод Гаусса-Ньютона

Получаем:

$$P(\theta^k + \Delta) = P(\theta^k) - SX\vec{\beta}\Delta + \frac{1}{2}\Delta^T(X^T S^2 X)\Delta$$

Медианная регрессия: метод Гаусса-Ньютона

Получаем:

$$P(\theta^k + \Delta) = P(\theta^k) - SX\vec{\beta}\Delta + \frac{1}{2}\Delta^T(X^T S^2 X)\Delta$$

Дифференцируем по Δ:

$$P'(\theta^k + \Delta) = -SX\vec{\beta} + (X^T S^2 X)\Delta = S^2 X\vec{r} + (X^T S^2 X)\Delta$$

Медианная регрессия: метод Гаусса-Ньютона

Получаем:

$$P(\theta^k + \Delta) = P(\theta^k) - SX\vec{\beta}\Delta + \frac{1}{2}\Delta^T(X^T S^2 X)\Delta$$

Дифференцируем по Δ:

$$P'(\theta^k + \Delta) = -SX\vec{\beta} + (X^T S^2 X)\Delta = S^2 X\vec{r} + (X^T S^2 X)\Delta$$

Таким образом, вектор

$$(X^T S^2 X)^{-1} S^2 X\vec{r}$$

Является точным решением задачи.

Медианная регрессия: метод Гаусса-Ньютона

Алгоритм:

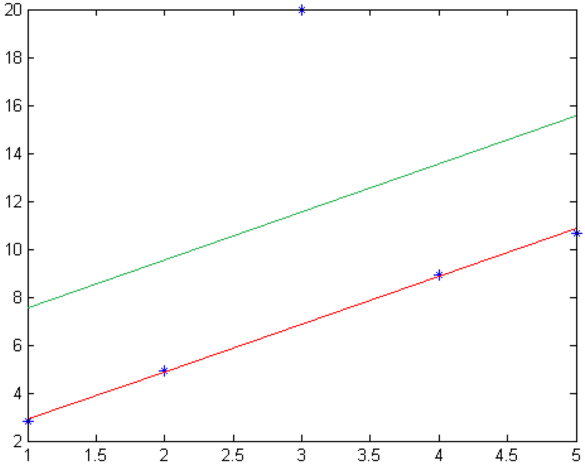
1. Выбрать ε таким, что $-2\varepsilon\ln(\varepsilon)$ было достаточно мало
2. Итерационно применять до сходимости следующую формулу:

$$\theta^{k+1} = \theta^k + (X^T S^2 X)^{-1} X^T S^2 y,$$

где

$$X = (x_1, ..., x_n), S^2 = diag(s_1^2, ..., s_n^2), s_i^2 = 1/(\varepsilon + |r_i|)$$

Пример применения медианной регрессии



Спасибо за внимание