Машинное обучение в науках о Земле

Михаил Криницкий
krinitsky.ma@phystech.edu

к.т.н., н.с.
Институт океанологии РАН им. П.П. Ширшова

Лаборатория взаимодействия океана и атмосферы и мониторинга климатических изменений (ЛВОАМКИ)

Общий принцип обучения по прецедентам (оптимизация функции ошибки)

$x \in \mathbb{X}$ — объекты, objects

$y \in \mathbb{Y}$ — ответы, labels

$\mathcal{F}\colon \mathbb{X} \to \mathbb{Y}$ — искомая закономерность

$\mathcal{T}\colon \{x_i; y_i\}$ — «обучающая выборка» (прецеденты), train dataset

Найти: $\hat{\mathcal{F}}\colon \{x_i\} \to \{y_i\}$

<u>один из способов решения:</u>

$\mathcal{L}\big(\hat{\mathcal{F}}(x)\big)$ — функционал ошибки (эмпирического риска, потерь), Loss function

$\hat{y}_i = \hat{\mathcal{F}}(x_i) = f(\vec{p}, x_i)$ — функционально задаваемая зависимость. **Предположение исследователя о виде закономерности**. Иногда задается параметрически, $\vec{p}$ — вектор параметров.

$$\mathcal{L} = L(\vec{p}, \mathcal{T}) \text{ — функция ошибки}$$

$$\hat{p} = \underset{\mathbb{P}}{\operatorname{argmin}}\big(L(\vec{p}, \mathcal{T})\big)$$

$$\hat{\mathcal{F}} = f(\hat{p}, x)$$

Общий принцип обучения по прецедентам (оптимизация функции ошибки)

$x \in \mathbb{X}$ — объекты, objects

$y \in \mathbb{Y}$ — ответы, labels

$\mathcal{F}: \mathbb{X} \rightarrow \mathbb{Y}$ — искомая закономерность

$\mathcal{T}: \{x_i; y_i\}$ — «обучающая выборка» (прецеденты), train dataset

Найти: $\hat{\mathcal{F}}: \{x_i\} \rightarrow \{y_i\}$

<u>один из способов решения:</u>

$\mathcal{L}(\hat{\mathcal{F}}(x))$ — функционал ошибки (эмпирического риска, потерь), Loss function

$\hat{y}_i = \hat{\mathcal{F}}(x_i) = f(\vec{p}, x_i)$ — функционально задаваемая зависимость. **Предположение исследователя о виде закономерности**. Иногда задается параметрически, $\vec{p}$ — вектор параметров.

$\mathcal{L} = L(\vec{p}, \mathcal{T})$ — функция ошибки

$$\hat{p} = \underset{\mathbb{P}}{\operatorname{argmin}}\left(L(\vec{p}, \mathcal{T})\right)$$

$$\hat{\mathcal{F}} = f(\hat{p}, x)$$

Чем руководствоваться при выборе функции ошибки?
КАКИЕ бывают функции ошибки?!

Обучение по прецедентам:
вероятностная постановка
принцип максимального правдоподобия maximum likelihood estimation

$\overrightarrow{\boldsymbol{x}_i}$ - признаковое описание объектов
$\overrightarrow{\boldsymbol{y}_i}$ - признаковое описание ответов
$p(\vec{\boldsymbol{x}}, \vec{\boldsymbol{y}})$ – (искомая, аппроксимируемая) **совместная плотность распределения событий** на множестве $X \times Y$

$\mathcal{T}$: $\{\overrightarrow{\boldsymbol{x}_i}; \overrightarrow{\boldsymbol{y}_i}\}$ — «обучающая выборка» (прецеденты), train dataset

Предположение!
(x_i, y_i) – выбираются из $\boldsymbol{p}(\boldsymbol{x}, \boldsymbol{y})$ независимо и случайно

Обучение по прецедентам:
вероятностная постановка
принцип максимального правдоподобия maximum likelihood estimation

$\boldsymbol{x_i}$ - признаковое описание объектов

$\boldsymbol{y_i}$ - признаковое описание ответов

$p(\boldsymbol{x}, \boldsymbol{y})$ – (искомая, аппроксимируемая) совместная плотность распределения событий на множестве $X \times Y$

$\phi(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta})$ - модель плотности распределения, предлагаемая исследователем

$\mathcal{T}$: $\{\boldsymbol{x_i}; \boldsymbol{y_i}\}$ — «обучающая выборка» (прецеденты), train dataset

Предположение!

(x_i, y_i) – выбираются из $\boldsymbol{p}(\boldsymbol{x}, \boldsymbol{y})$ независимо и случайно

Обучение по прецедентам:
вероятностная постановка
принцип максимального правдоподобия maximum likelihood estimation

$\boldsymbol{x_i}$ - признаковое описание объектов

$\boldsymbol{y_i}$ - признаковое описание ответов

$p(\boldsymbol{x}, \boldsymbol{y})$ – (искомая, аппроксимируемая) совместная плотность распределения событий на множестве $X \times Y$

$\phi(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta})$ - модель плотности распределения, предлагаемая исследователем

$\mathcal{T}\colon \{\boldsymbol{x_i}; \boldsymbol{y_i}\}$ — «обучающая выборка» (прецеденты), train dataset

Предположение!

(x_i, y_i) – выбираются из $\boldsymbol{p}(\boldsymbol{x}, \boldsymbol{y})$ независимо и случайно

MLE

$\phi(\boldsymbol{x_i}, \boldsymbol{y_i}, \boldsymbol{\theta})$ - правдоподобие для одного экземпляра выборки

$L(\{\boldsymbol{x_i}\}, \{\boldsymbol{y_i}\}, \boldsymbol{\theta}) = \prod_{i=1}^{N} \phi(\boldsymbol{x_i}, \boldsymbol{y_i}, \boldsymbol{\theta})$ - правдоподобие выборки

$$\theta^* = \underset{\Theta}{\operatorname{argmax}}\, L(\{\boldsymbol{x_i}\}, \{\boldsymbol{y_i}\}, \boldsymbol{\theta})$$

Функция потерь определяется видом модели плотности распределения $\phi(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta})$, предложенной исследователем!

Правдоподобие выборки $L(\boldsymbol{\theta}, \mathcal{T})$ – **максимизировать** (в пространстве параметров Θ)

Функцию потерь $\mathcal{L}(\boldsymbol{\theta}, \mathcal{T})$ – **минимизировать** (в пространстве параметров Θ)

Обучение по прецедентам. Вероятностная постановка, MLE

Примеры

Линейная регрессия

MSE

$$\phi(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta}) = \theta x + \epsilon,$$

$$p(\epsilon) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\epsilon^2}{2\sigma^2}}$$

=> $$\mathcal{L}(\boldsymbol{\theta}, \mathcal{T}) = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{y_i} - \boldsymbol{\theta x_i})^2$$

MAE

$$\phi(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta}) = \theta x + \epsilon,$$

$$p(\epsilon) = \frac{1}{2b} e^{-\frac{|\epsilon|}{b}}$$

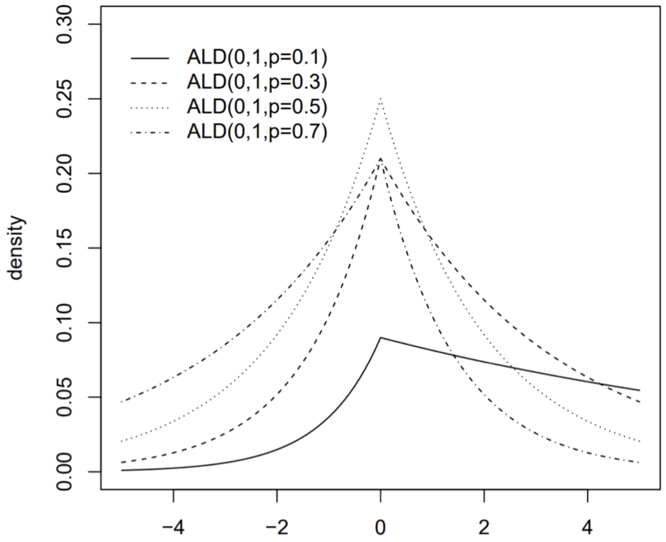
=> $$\mathcal{L}(\boldsymbol{\theta}, \mathcal{T}) = \frac{1}{N}\sum_{i=1}^{N}|\boldsymbol{y_i} - \boldsymbol{\theta x_i}|$$

~~previously on ML4ES~~

Квантильная регрессия

Bera, Anil & Galvao, Antonio & Montes-Rojas, Gabriel & Park, Sung Y.. (2015). **Asymmetric Laplace Regression: Maximum Likelihood, Maximum Entropy and Quantile Regression**. Journal of Econometric Methods. 10.1515/jem-2014-0018.

Sánchez, B. L., Lachos, H. V., & Labra, V. F. (2013). **Likelihood based inference for quantile regression using the asymmetric Laplace distribution**. Journal of Statistical Computation and Simulation, 81, 1565-1578.



Asymmetric Laplace density. From Sánchez et al.

Обучение по прецедентам. Вероятностная постановка, MLE
Примеры

Логистическая регрессия (logistic regression)

Coming soon!
(its for classification problem)

Решение задач типа ОБУЧЕНИЕ С УЧИТЕЛЕМ

Михаил Криницкий
krinitsky.ma@phystech.edu

к.т.н., н.с.
Институт океанологии РАН им. П.П. Ширшова

Лаборатория взаимодействия океана и атмосферы и мониторинга климатических изменений (ЛВОАМКИ)

ОБЩАЯ СХЕМА РЕШЕНИЯ ЗАДАЧ ОБУЧЕНИЯ С УЧИТЕЛЕМ

обучаем (тренируем) модель **на имеющихся данных**

«Обучение»

Объекты

признаки

объекты

	1	2	3	4	5	6	7
1	-0.0685	-0.0676	-0.0643	-0.0617	-0.0550	-0.0488	-0.0365
2	-0.0706	-0.0726	-0.0633	-0.0579	-0.0490	-0.0390	-0.0267
3	-0.0691	-0.0766	-0.0676	-0.0589	-0.0484	-0.0350	-0.0172
4	-0.0631	-0.0719	-0.0662	-0.0557	-0.0441	-0.0317	-0.0144
5	-0.0500	-0.0634	-0.0603	-0.0502	-0.0384	-0.0263	-0.0115
6	-0.0377	-0.0491	-0.0449	-0.0352	-0.0214	-0.0062	0.0056
7	-0.0375	-0.0319	-0.0235	-0.0133	0.0012	0.0169	0.0282
8	-0.0540	-0.0310	-0.0169	6.9177e-...	0.0181	0.0339	0.0445
9	-0.0725	-0.0550	-0.0158	0.0145	0.0323	0.0501	0.0641
10	-0.0789	-0.0844	-0.0512	-0.0010	0.0387	0.0612	0.0810
11	-0.0803	-0.0965	-0.0789	-0.0543	-0.0036	0.0405	0.0685
12	-0.0995	-0.1248	-0.1134	-0.0950	-0.0699	-0.0281	0.0261
13	-0.1026	-0.1306	-0.1283	-0.1171	-0.0982	-0.0806	-0.0412
14	-0.0977	-0.1284	-0.1301	-0.1188	-0.1054	-0.0932	-0.0677
15	-0.0840	-0.1133	-0.1138	-0.1040	-0.0895	-0.0750	-0.0511
16	-0.0820	-0.1089	-0.1072	-0.0967	-0.0814	-0.0669	-0.0464
17	-0.0655	-0.0685	-0.0617	-0.0595	-0.0549	-0.0499	-0.0396

Целевая переменная

	1
1	-0.0685
2	-0.0706
3	-0.0691
4	-0.0631
5	-0.0500
6	-0.0377
7	-0.0375
8	-0.0540
9	-0.0725
10	-0.0789
11	-0.0803
12	-0.0995
13	-0.1026
14	-0.0977
15	-0.0840
16	-0.0820
17	-0.0655

Модель - метод, структура, значения параметров

$$\hat{y}_i = f(\vec{p}, x_i)$$

Настройка параметров модели

НАСТРОЕННАЯ МОДЕЛЬ

Режим использования

новые объекты

	1	2	3	4	5	6	7	8	9
1	-0.0685	-0.0676	-0.0643	-0.0617	-0.0550	-0.0488	-0.0365	-0.0235	-0.0133
2	-0.0706	-0.0726	-0.0633	-0.0579	-0.0490	-0.0390	-0.0267	-0.0136	-0.0044
3	-0.0691	-0.0766	-0.0676	-0.0589	-0.0484	-0.0350	-0.0172	-0.0038	0.0054
4	-0.0631	-0.0719	-0.0662	-0.0557	-0.0441	-0.0317	-0.0144	-0.0022	0.0084
5	-0.0500	-0.0634	-0.0603	-0.0502	-0.0384	-0.0263	-0.0115	0.0020	0.0152
6	-0.0377	-0.0491	-0.0449	-0.0352	-0.0214	-0.0062	0.0056	0.0145	0.0238
7	-0.0375	-0.0319	-0.0235	-0.0133	0.0012	0.0169	0.0282	0.0355	0.0403

НАСТРОЕННАЯ МОДЕЛЬ

Результат: значения целевой переменной

…
Вечор, ты помнишь, вьюга злилась,
На мутном небе мгла носилась;
Луна, как бледное пятно,
Сквозь тучи мрачные желтела,
И ты печальная сидела —
А нынче... погляди в окно:

Под голубыми небесами
Великолепными коврами,
Блестя на солнце, снег лежит;
Прозрачный лес один чернеет,
И ель сквозь иней зеленеет,
И речка подо льдом блестит.
…

А.С. Пушкин, «Зимнее утро»

…
Буря мглою небо кроет,
Вихри снежные крутя;
То, как зверь, она завоет,
То заплачет, как дитя.
Выпьем, добрая подружка
Бедной юности моей,
Выпьем с горя; где же кружка?
Сердцу будет веселей.

А.С. Пушкин, «Зимний вечер»

…
Было так: Нева, как зверь, стонала,
Серые ломая гребешки,
Колыхались барки у причала,
И царапал стынущие щеки
Острый дождь, ложась, как плащ широкий,
Над гранитным логовом реки.
…

В. Рождественский, «Октябрьская погода»

О признаковом описании событий (объектов) в геофизике

…
Вечор, ты помнишь, вьюга злилась,
На мутном небе мгла носилась;
Луна, как бледное пятно,
Сквозь тучи мрачные желтела,
И ты печальная сидела —
А нынче... погляди в окно:

Под голубыми небесами
Великолепными коврами,
Блестя на солнце, снег лежит;
Прозрачный лес один чернеет,
И ель сквозь иней зеленеет,
И речка подо льдом блестит.
…

А.С. Пушкин, «Зимнее утро»

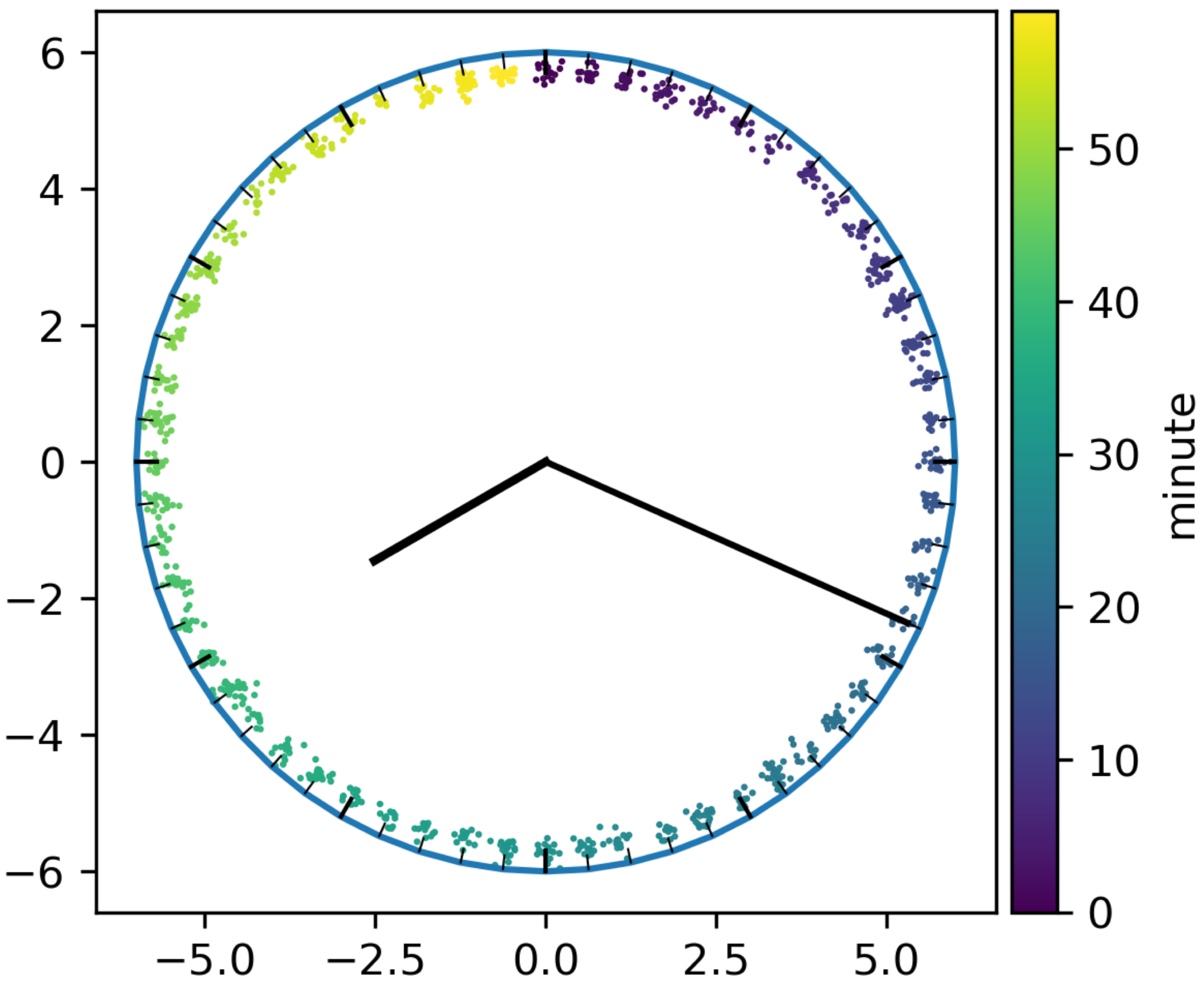
…
Буря мглою небо кроет,
Вихри снежные крутя;
То, как зверь, она завоет,
То заплачет, как дитя.
Выпьем, добрая подружка
Бедной юности моей,
Выпьем с горя; где же кружка?
Сердцу будет веселей.

А.С. Пушкин, «Зимний вечер»

…
Было так: Нева, как зверь, стонала,
Серые ломая гребешки,
Колыхались барки у причала,
И царапал стынущие щеки
Острый дождь, ложась, как плащ широкий,
Над гранитным логовом реки.
…

В. Рождественский, «Октябрьская погода»

Решение задачи восстановления регрессии: ПРИМЕР

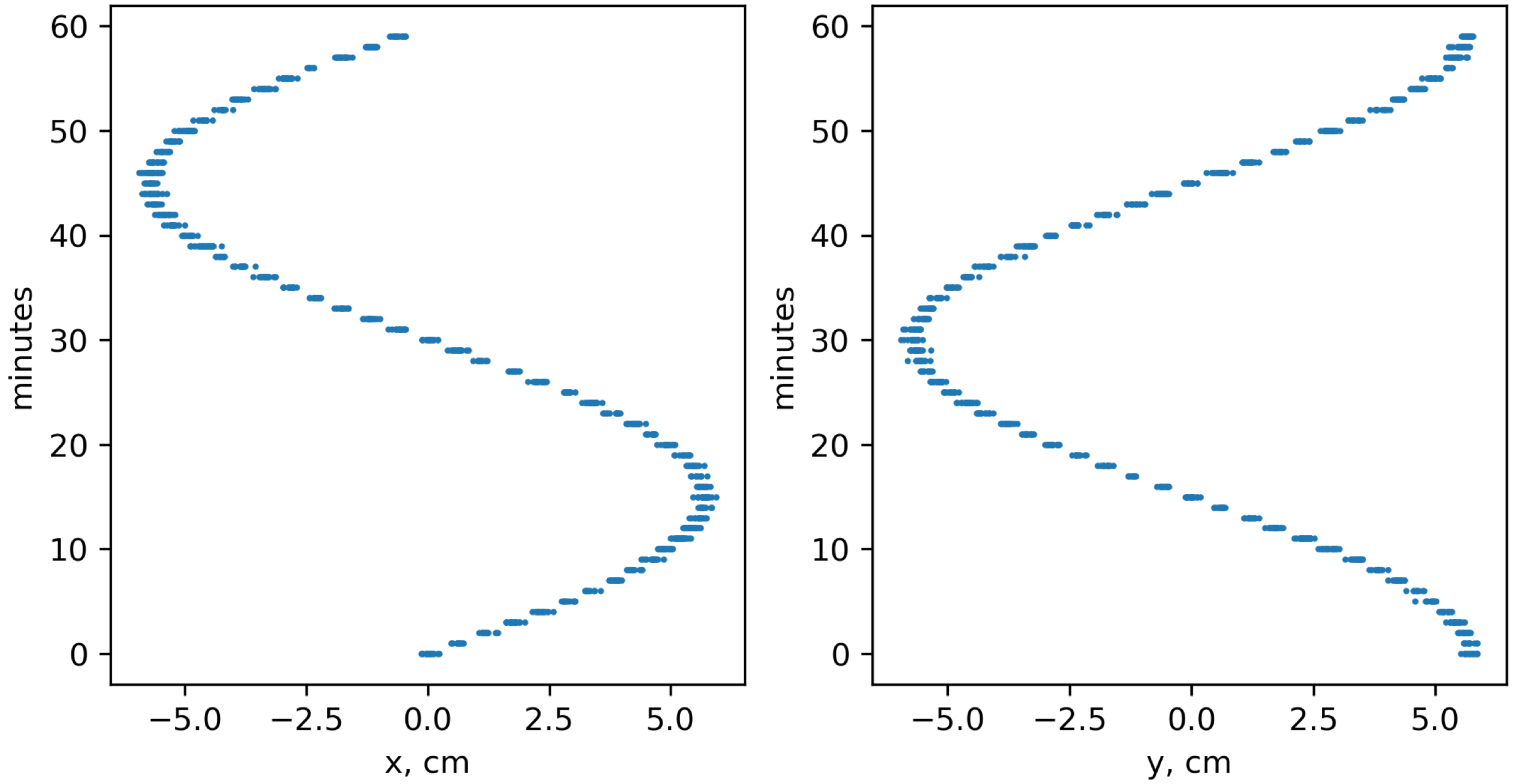


Синтетическая задача, “toy problem”

События x_i: наблюдения циферблата часов
Признаковое описание событий $\overrightarrow{x_i}$: координаты конца минутной стрелки
Целевая переменная m_i: минутная компонента времени

Решение задачи восстановления регрессии: ПРИМЕР

Исследование данных: визуализация, поиск структуры



Решение задачи восстановления регрессии: ПРИМЕР

Построение и настройка модели

Возьмем очень слабую модель

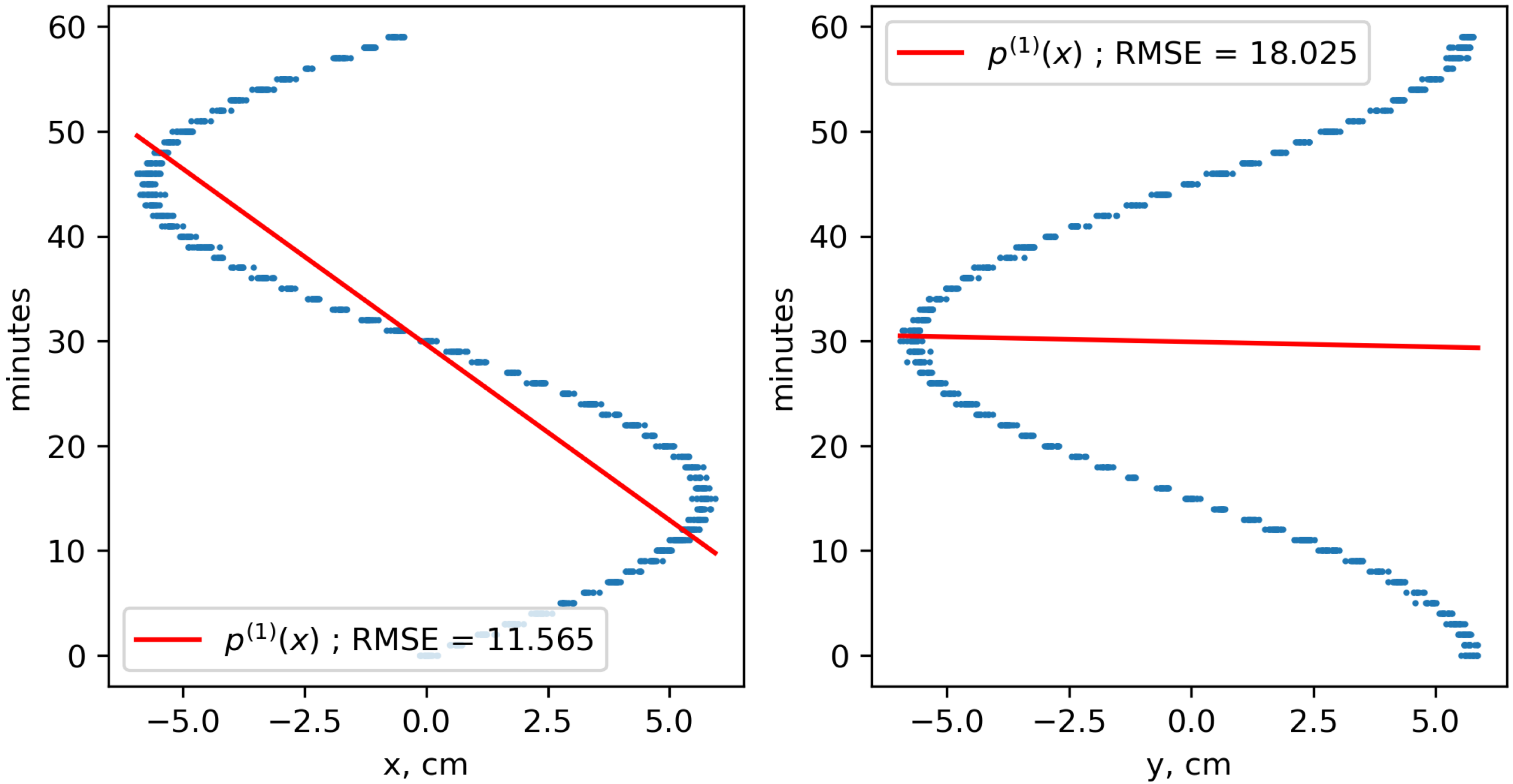
Модель в <u>задаче восстановления регрессии</u>: $\boldsymbol{\hat{m}_i} = \boldsymbol{f}(\vec{\boldsymbol{p}}, \boldsymbol{x_i}) = \boldsymbol{k}\boldsymbol{x_i} + \boldsymbol{b}$

Функция потерь: $\mathcal{L}(\vec{\boldsymbol{p}}, \{\boldsymbol{x_i}\}, \{\boldsymbol{m_i}\}) = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{f}(\vec{\boldsymbol{p}}, \boldsymbol{x_i}) - \boldsymbol{m_i})^2$

Решение (оценка параметров $\vec{\boldsymbol{p}}$): $\overrightarrow{p^*} = \underset{\mathbb{P}}{\operatorname{argmin}}(\mathcal{L}(\vec{\boldsymbol{p}}, \{\boldsymbol{x_i}\}, \{\boldsymbol{m_i}\}))$

Решение задачи восстановления регрессии: ПРИМЕР

Результаты модели



Решение задачи восстановления регрессии: ПРИМЕР

Построение и настройка модели

Возьмем модель посильнее

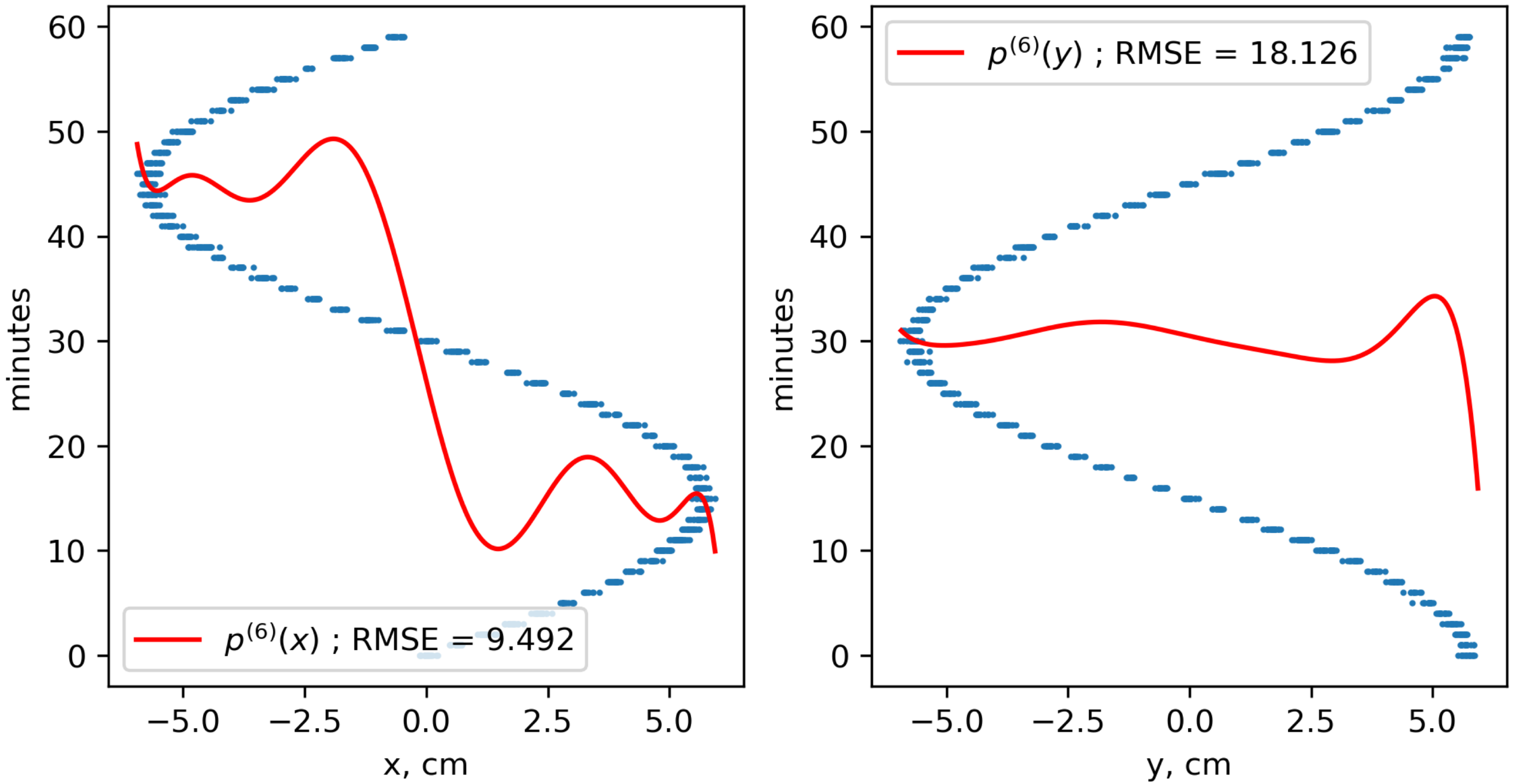
Модель в задаче восстановления регрессии: $\widehat{\boldsymbol{m}}_i = \boldsymbol{f}(\vec{\boldsymbol{p}}, \boldsymbol{x}_i) = \boldsymbol{poly}^{(6)}(\boldsymbol{x}_i)$

Функция потерь: $\mathcal{L}(\vec{\boldsymbol{p}}, \{\boldsymbol{x}_i\}, \{\boldsymbol{m}_i\}) = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{f}(\vec{\boldsymbol{p}}, \boldsymbol{x}_i) - \boldsymbol{m}_i)^2$

Решение (оценка параметров $\vec{\boldsymbol{p}}$): $\overrightarrow{p^*} = \underset{\mathbb{P}}{\operatorname{argmin}}(\mathcal{L}(\vec{\boldsymbol{p}}, \{\boldsymbol{x}_i\}, \{\boldsymbol{m}_i\}))$

Решение задачи восстановления регрессии: ПРИМЕР

Результаты модели



Решение задачи восстановления регрессии: ПРИМЕР

Построение и настройка модели

Возьмем нейросеть

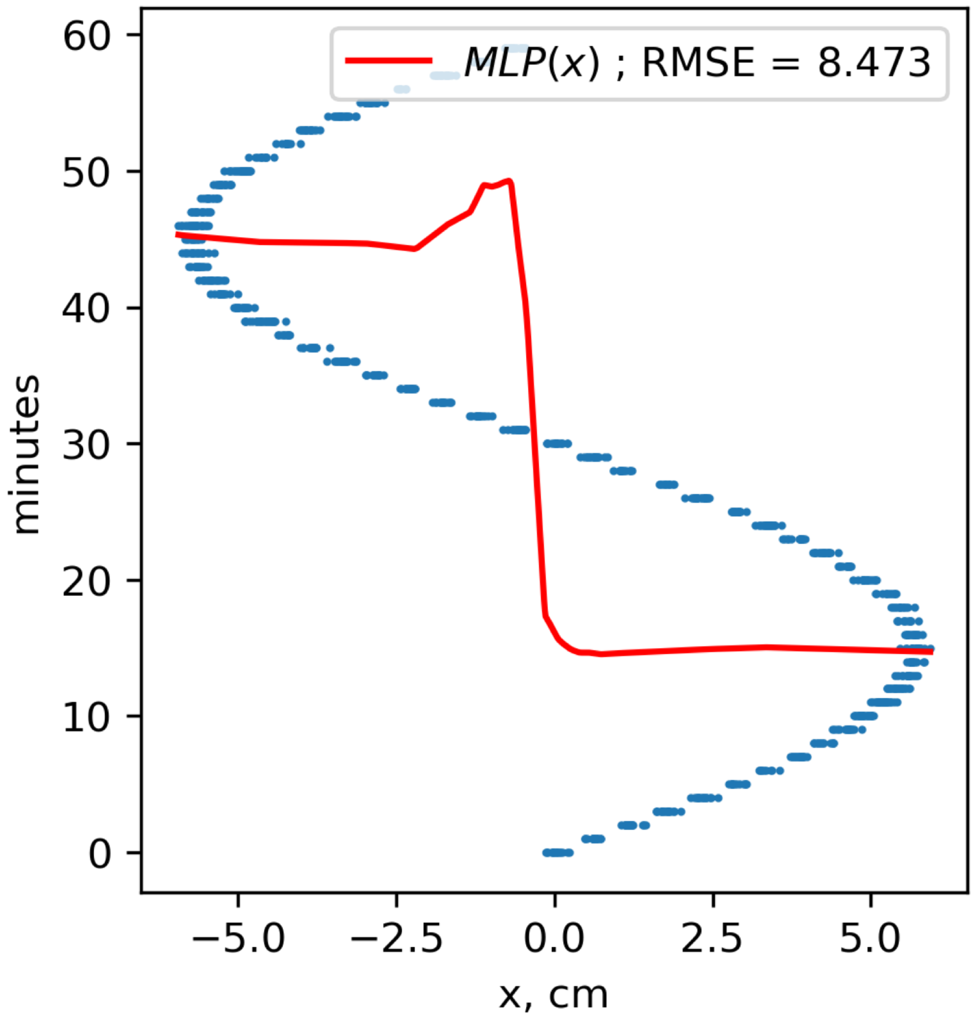
$$\boldsymbol{\widehat{m}_i} = \boldsymbol{MLP}(\vec{\boldsymbol{p}}, \boldsymbol{x_i})$$

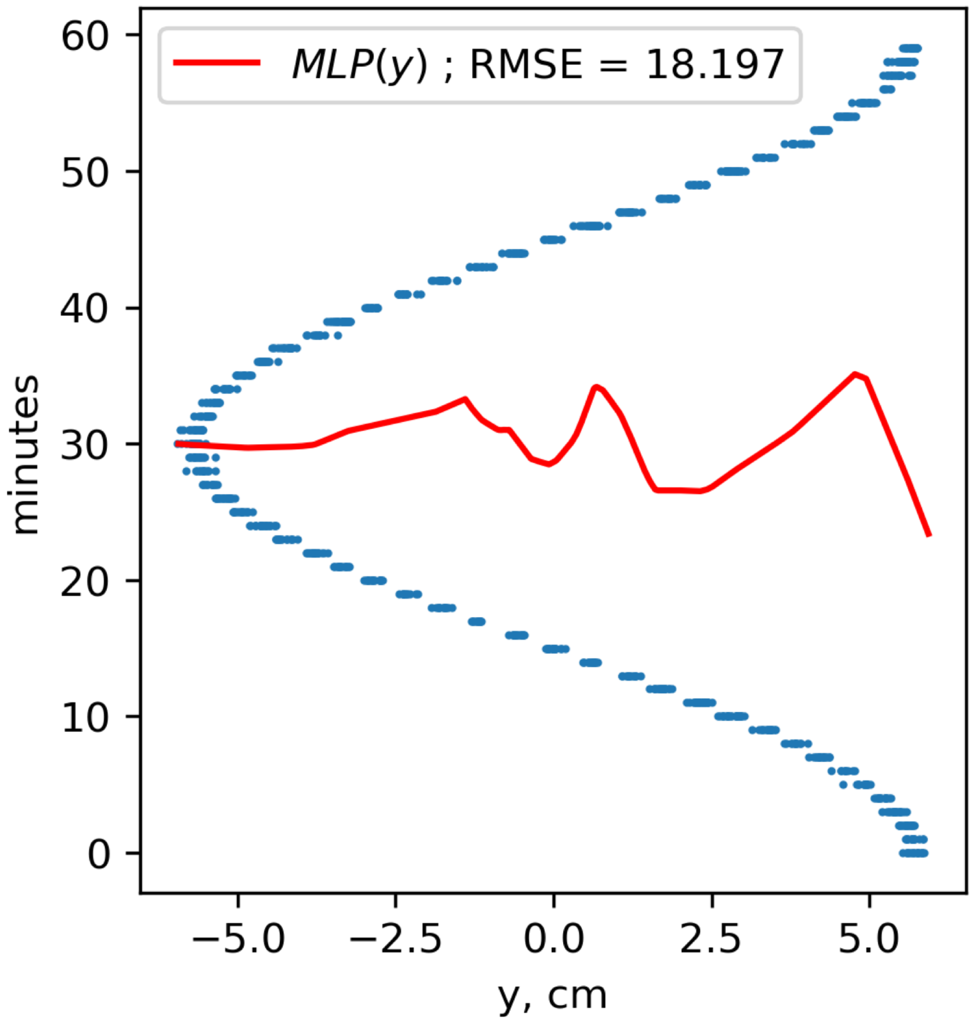
$$\mathcal{L}(\vec{\boldsymbol{p}}, \{\boldsymbol{x_i}\}, \{\boldsymbol{m_i}\}) = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{MLP}(\vec{\boldsymbol{p}}, \boldsymbol{x_i}) - \boldsymbol{m_i})^2$$

$$\overrightarrow{p^*} = \underset{\mathbb{P}}{\operatorname{argmin}}(\mathcal{L}(\vec{\boldsymbol{p}}, \{\boldsymbol{x_i}\}, \{\boldsymbol{m_i}\}))$$

Решение задачи восстановления регрессии: ПРИМЕР

Результаты модели





Решение задачи восстановления регрессии: ПРИМЕР

ЧТО ЖЕ ДЕЛАТЬ?

что-то не так с постановкой задачи? (не тот тип задачи? не та целевая переменная?)

что-то не так с признаковым описанием событий? (нерелевантное? неполное? шумное?)

что-то не так с разметкой? (шумная? некорректная? много? мало?)

что-то не так с моделью? (слишком простая? слишком сложная? не подходит для этой задачи?)

что-то не так с программным кодом?

что-то не так с исследователем?

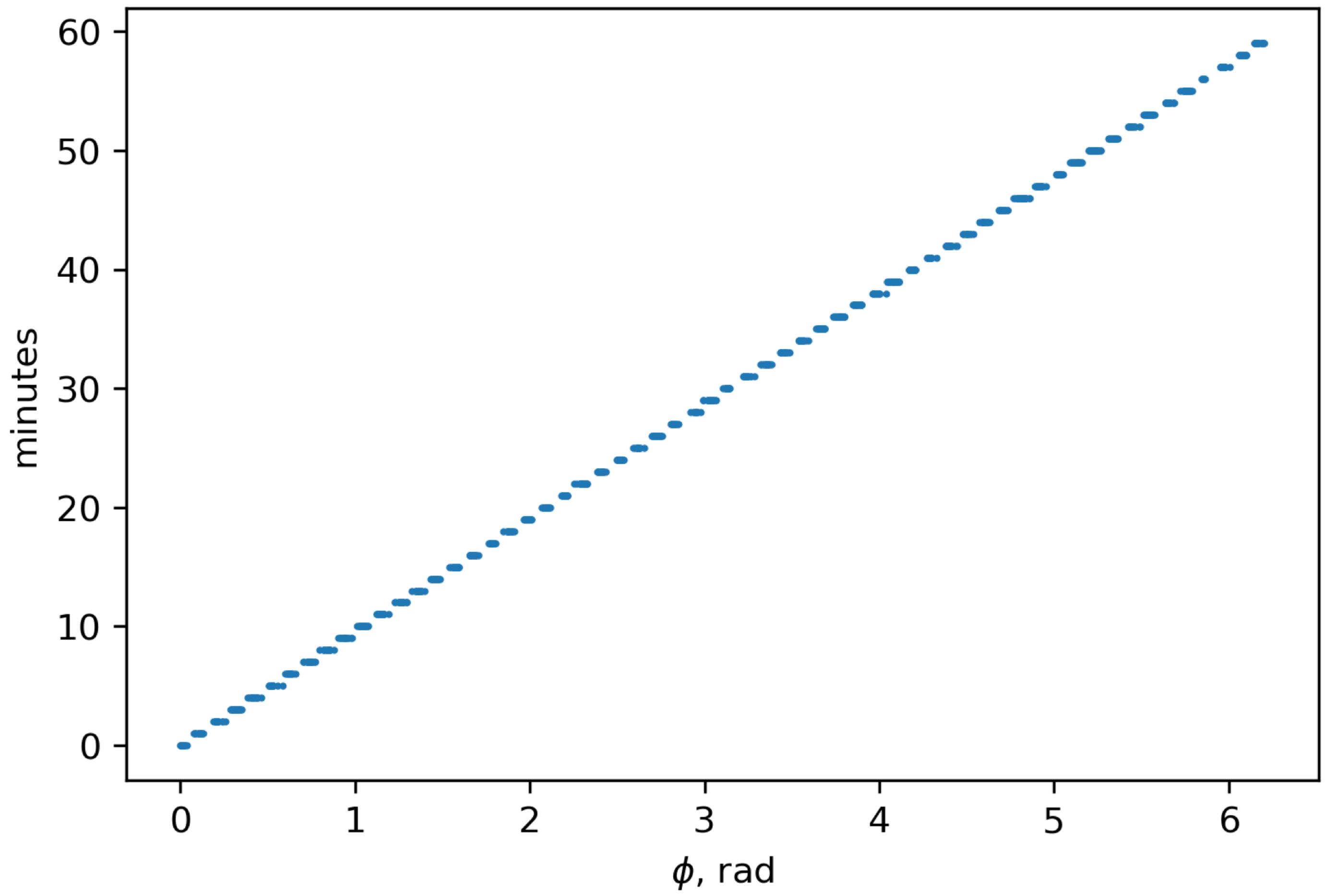
МОЖЕТ, ПРОСТО НЕТ ЗАКОНОМЕРНОСТИ?

Решение задачи восстановления регрессии: ПРИМЕР

изобретать (более информативные) признаки

Новое признаковое описание событий: $\vec{x_i}$ - угол отклонения минутной стрелки

$$x_i = \phi = \begin{cases} \arccos\left(\frac{x}{\sqrt{x^2+y^2}}\right), \text{если } x \geq 0 \\ 2\pi - \arccos\left(\frac{x}{\sqrt{x^2+y^2}}\right), \text{если } x < 0 \end{cases}$$



Решение задачи восстановления регрессии: ПРИМЕР

Построение и настройка модели

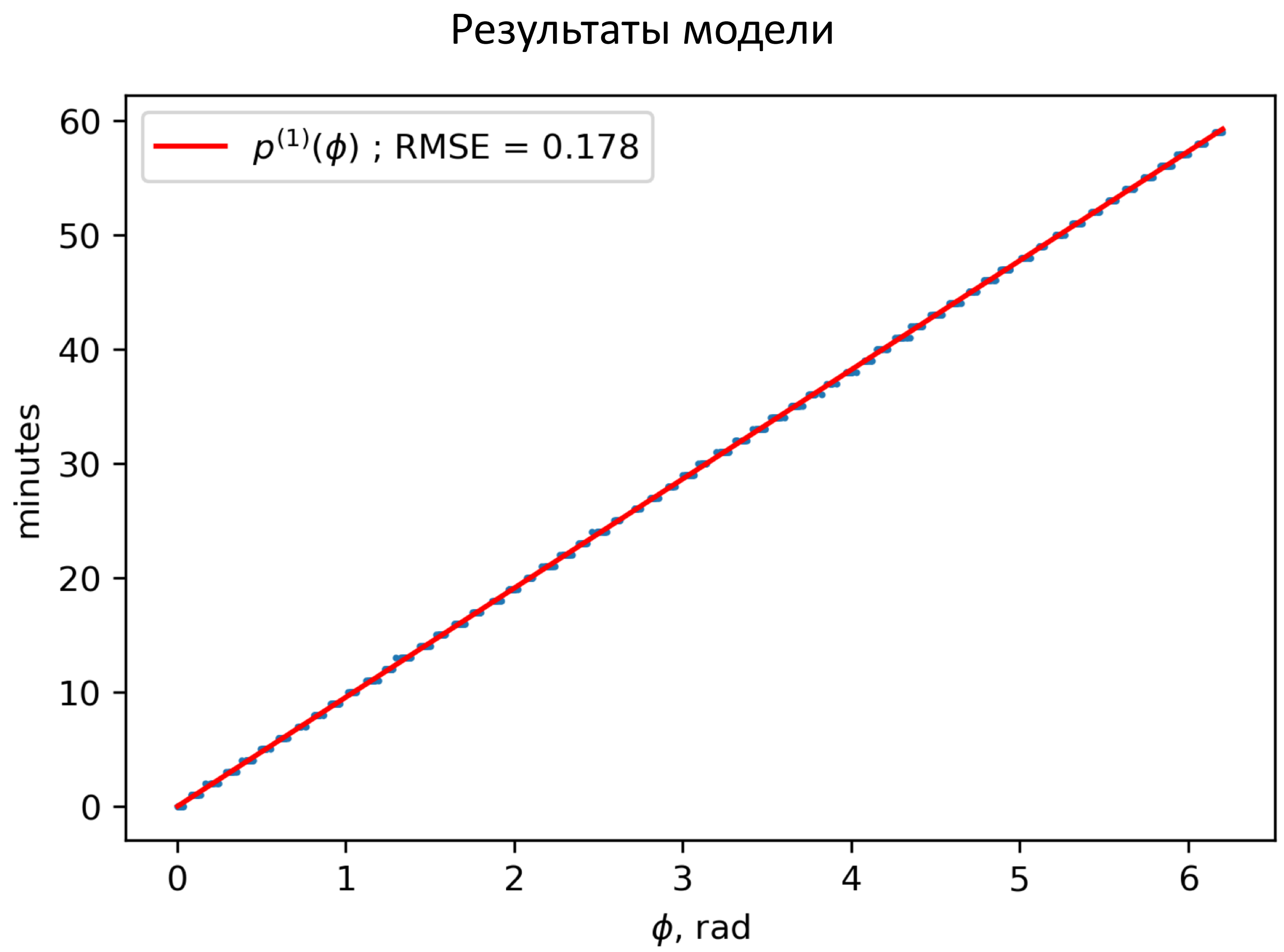
Возьмем очень слабую модель

Модель в задаче восстановления регрессии: $\widehat{\boldsymbol{m}}_i = \boldsymbol{f}(\vec{\boldsymbol{p}}, \boldsymbol{\phi}_i) = \boldsymbol{k}\boldsymbol{\phi}_i + \boldsymbol{b}$

Функция потерь: $\mathcal{L}(\vec{\boldsymbol{p}}, \{\boldsymbol{\phi}_i\}, \{\boldsymbol{m}_i\}) = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{f}(\vec{\boldsymbol{p}}, \boldsymbol{\phi}_i) - \boldsymbol{m}_i)^2$

Решение (оценка параметров $\vec{\boldsymbol{p}}$): $\overrightarrow{p^*} = \underset{\mathbb{P}}{\operatorname{argmin}}(\mathcal{L}(\vec{\boldsymbol{p}}, \{\boldsymbol{\phi}_i\}, \{\boldsymbol{m}_i\}))$

Решение задачи восстановления регрессии: ПРИМЕР



Решение задачи восстановления регрессии: ПРИМЕР

использовать (более) полную информацию о событиях

Новое признаковое описание событий: $\overrightarrow{\boldsymbol{x_i}}$ - **обе** координаты x, y конца минутной стрелки

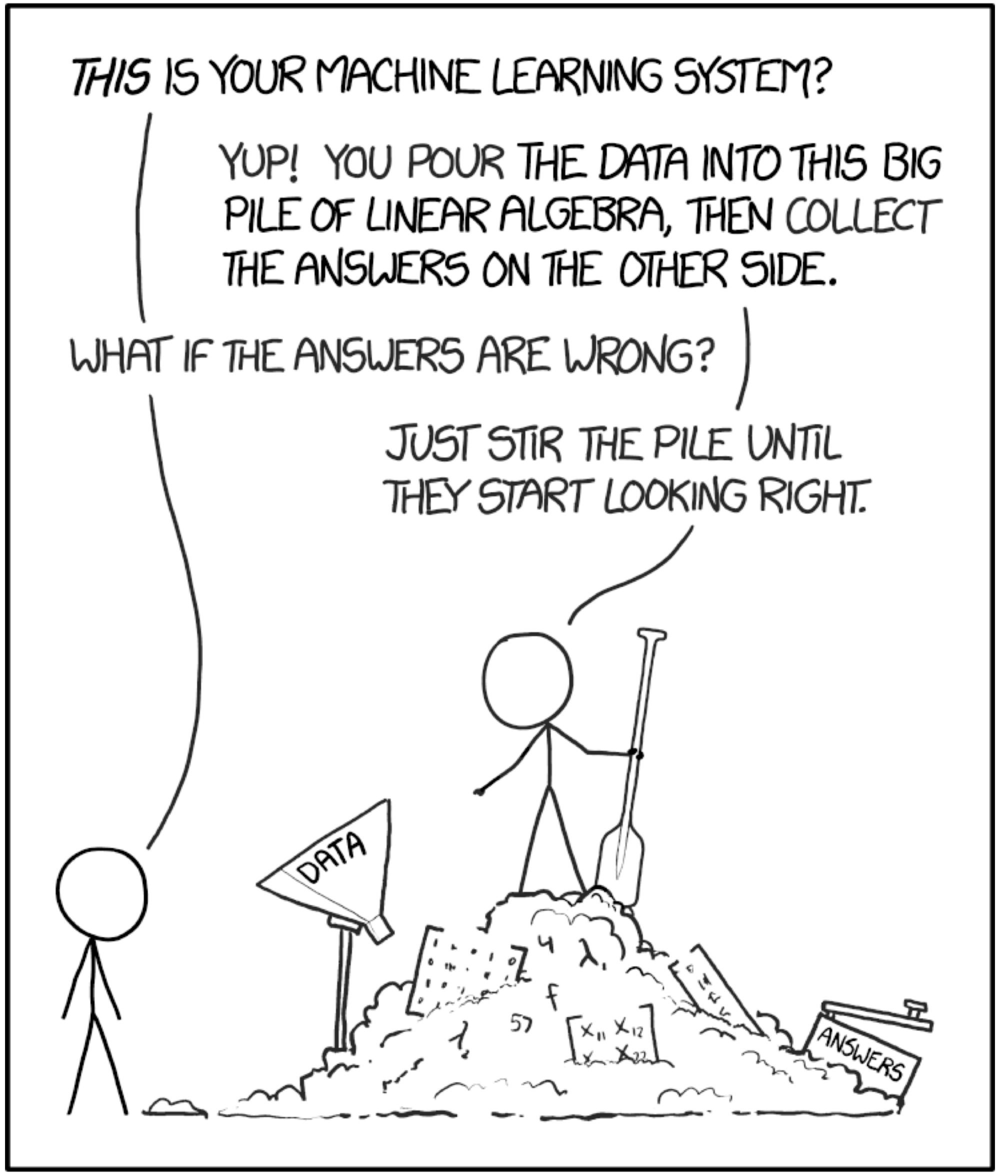
Возьмем нейросеть

$$\widehat{\boldsymbol{m}}_{\boldsymbol{i}} = \boldsymbol{MLP}(\vec{\boldsymbol{p}}, \overrightarrow{\boldsymbol{x_i}})$$

$$\mathcal{L}(\vec{\boldsymbol{p}}, \{\overrightarrow{\boldsymbol{x_i}}\}, \{\boldsymbol{m_i}\}) = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{MLP}(\vec{\boldsymbol{p}}, \overrightarrow{\boldsymbol{x_i}}) - \boldsymbol{m_i})^2$$

$$\overrightarrow{p^*} = \underset{\mathbb{P}}{\operatorname{argmin}}(\mathcal{L}(\vec{\boldsymbol{p}}, \{\overrightarrow{\boldsymbol{x_i}}\}, \{\boldsymbol{m_i}\}))$$

Качество модели: $RMSE = 0.28m$


THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS