



Машинное обучение в науках о Земле

Михаил Криницкий

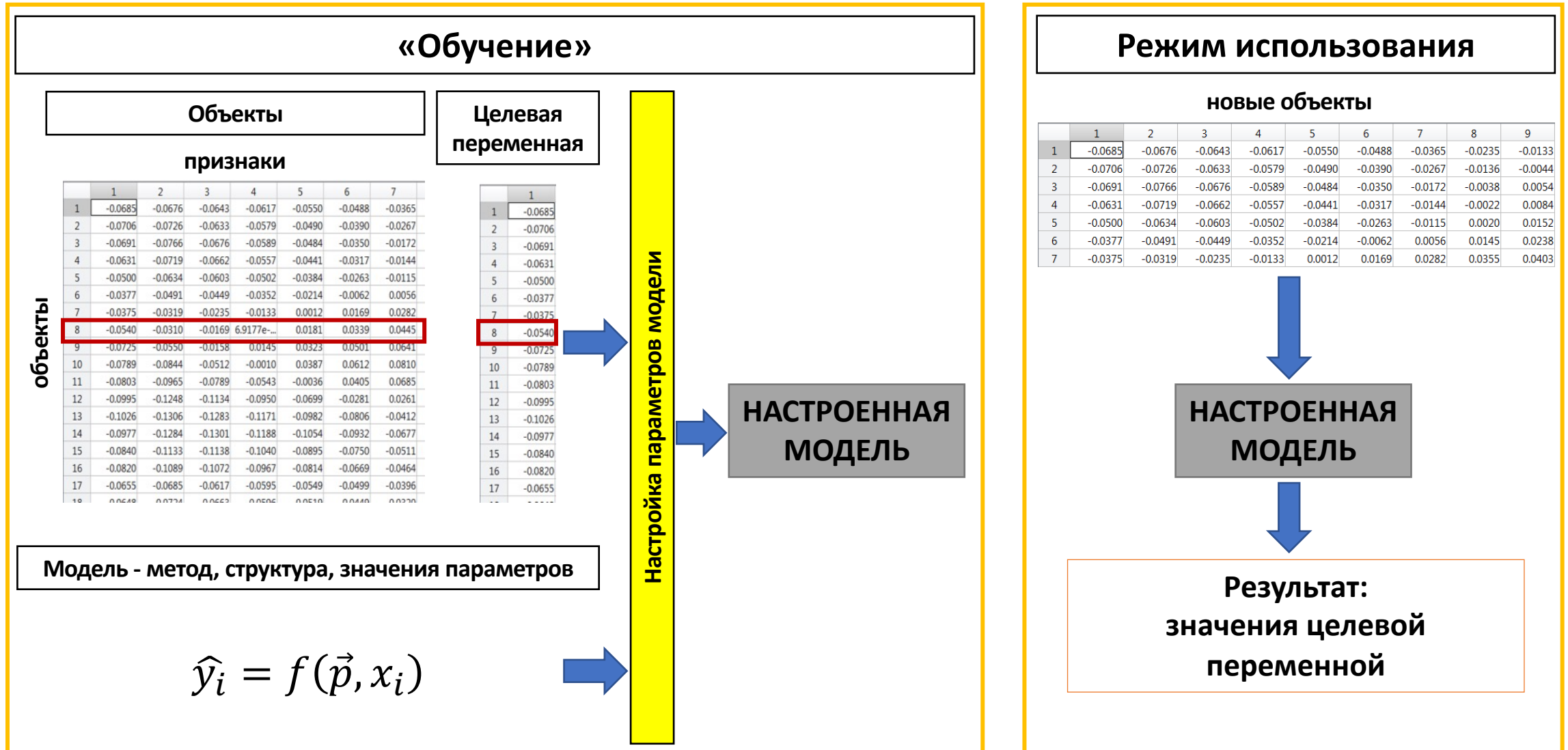
krinitsky.ma@phystech.edu

К.Т.Н.

Зав. лабораторией машинного обучения в науках о Земле МФТИ
с.н.с. Института океанологии РАН им. П.П. Ширшова

ОБЩАЯ СХЕМА РЕШЕНИЯ ЗАДАЧ ОБУЧЕНИЯ С УЧИТЕЛЕМ

обучаем (тренируем) модель на имеющихся данных



Пару слов о признаковом описании

...

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 стынувшие 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»

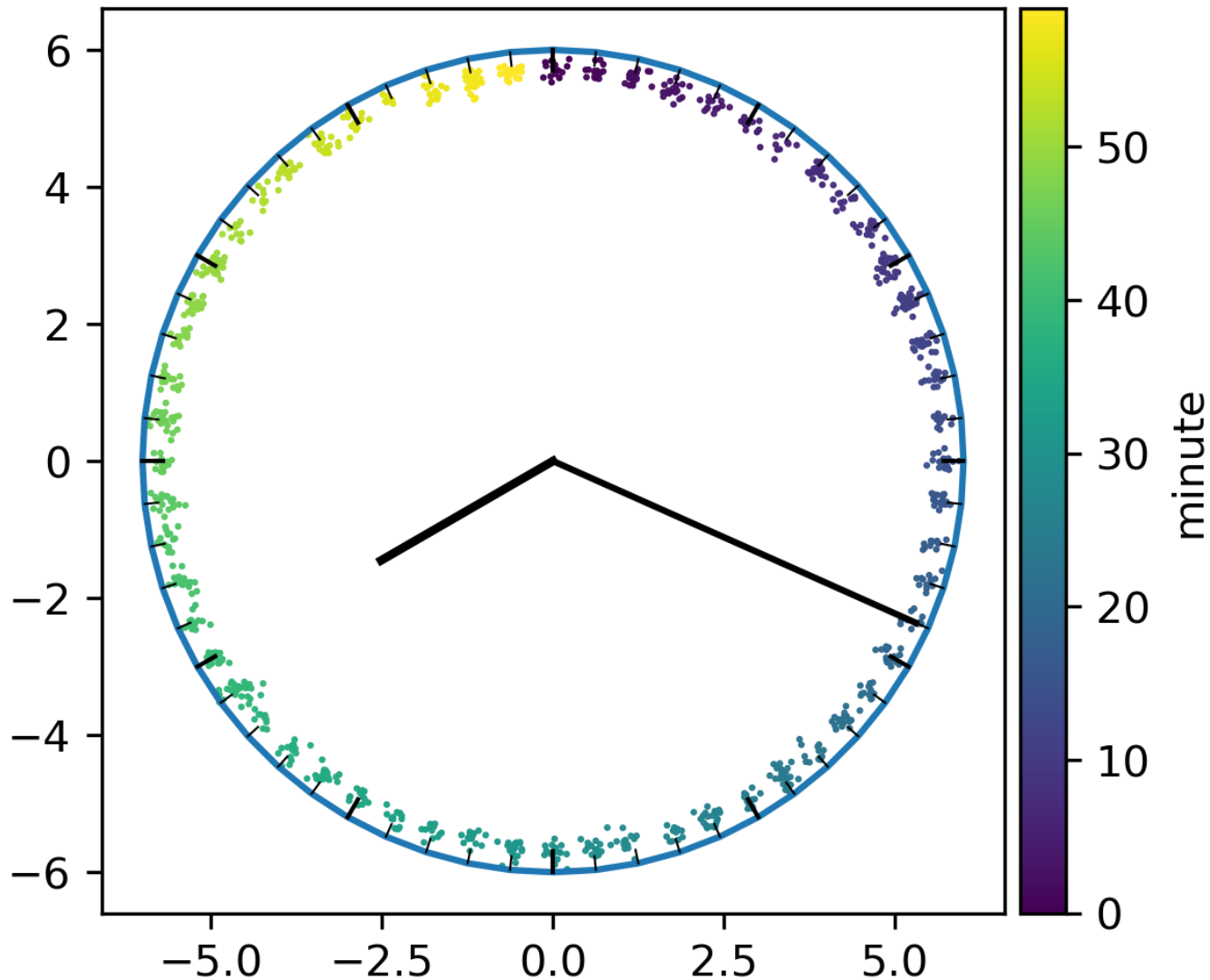
...

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

...

В. Рождественский, «Октябрьская погода»

Решение задачи восстановления регрессии: ПРИМЕР



Синтетическая задача, “toy problem”

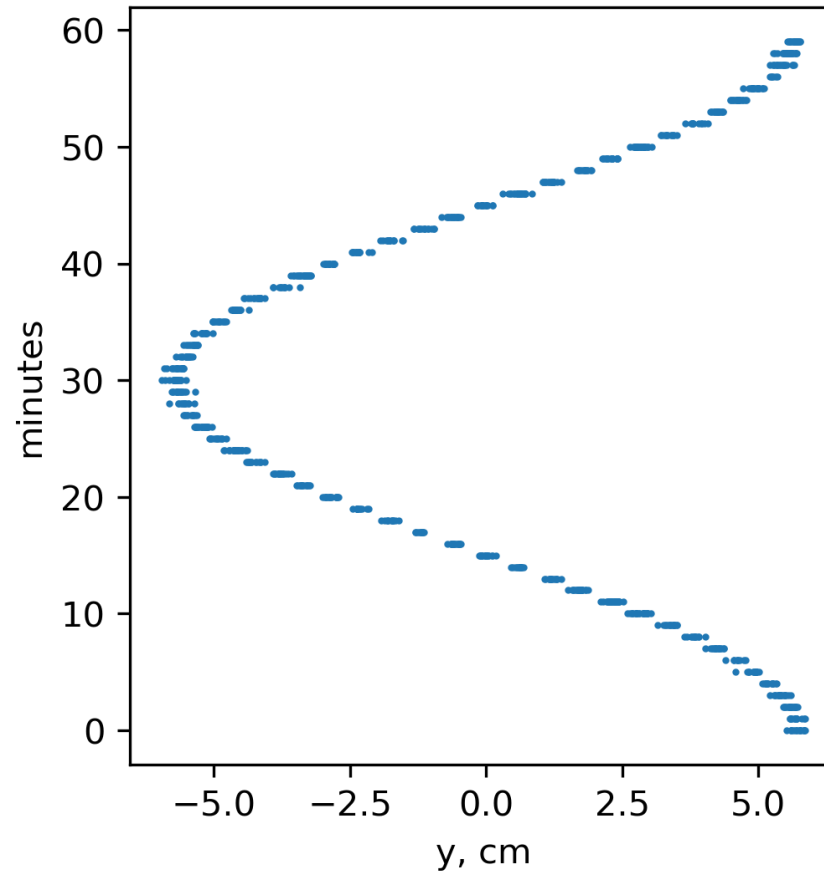
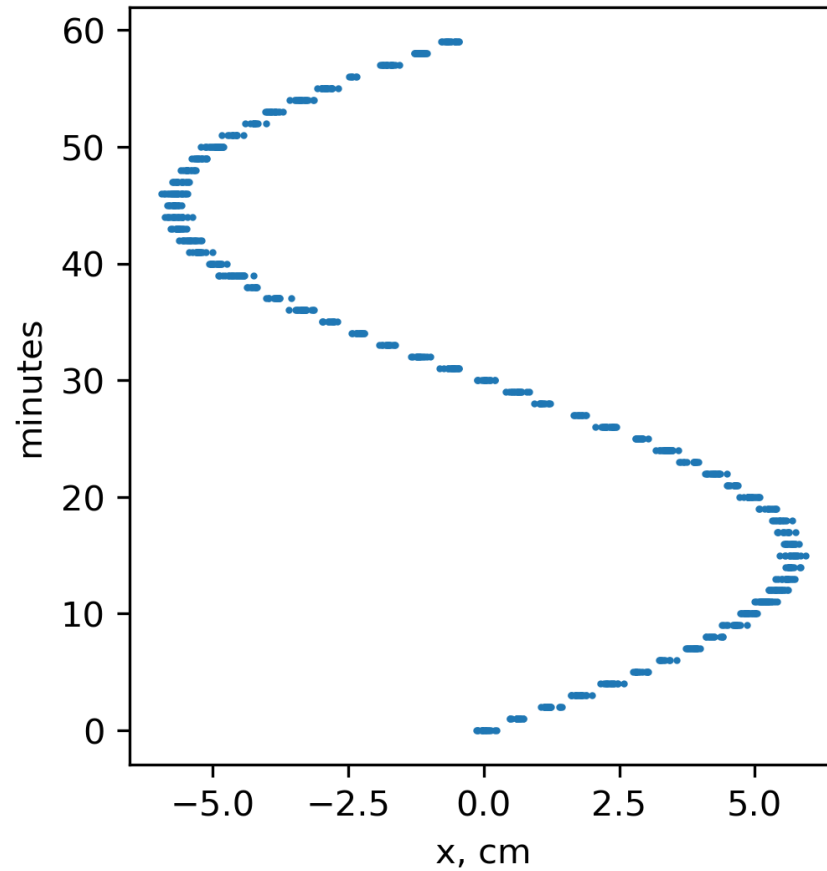
События x_i : наблюдения циферблата часов

Признаковое описание событий \vec{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

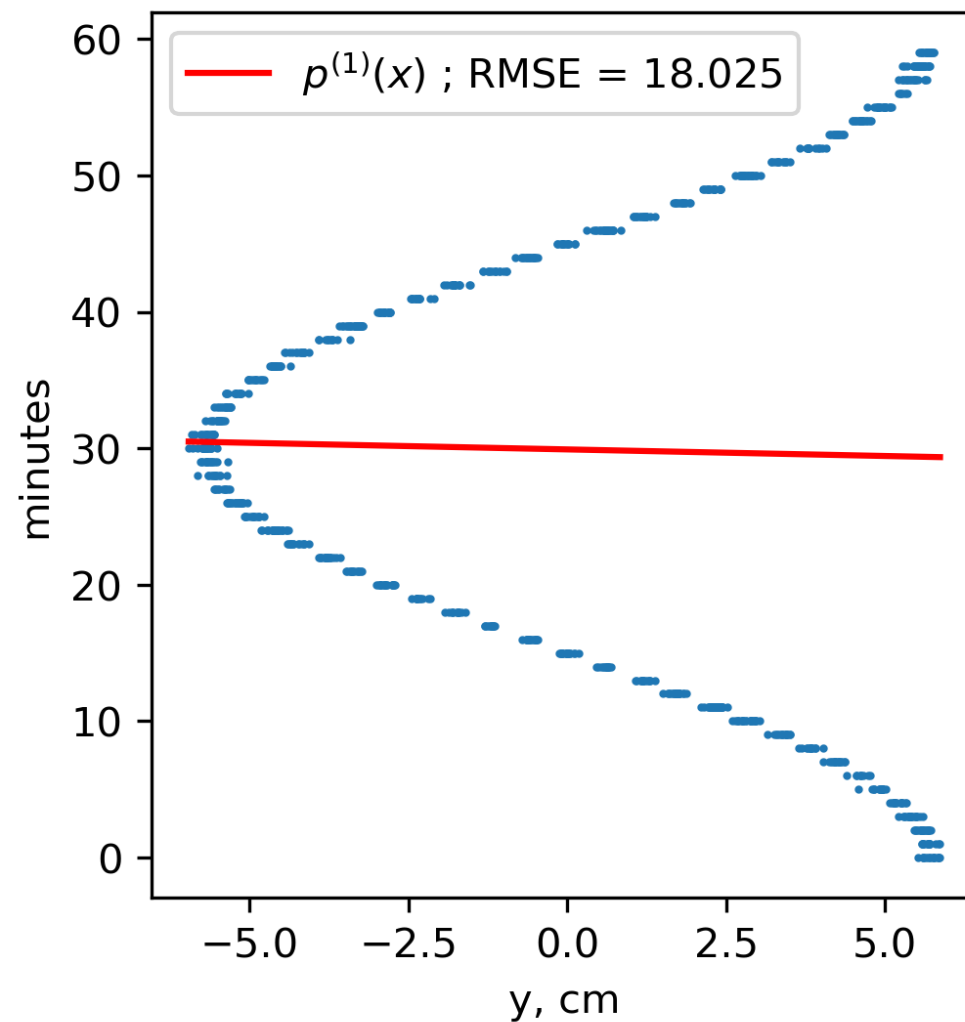
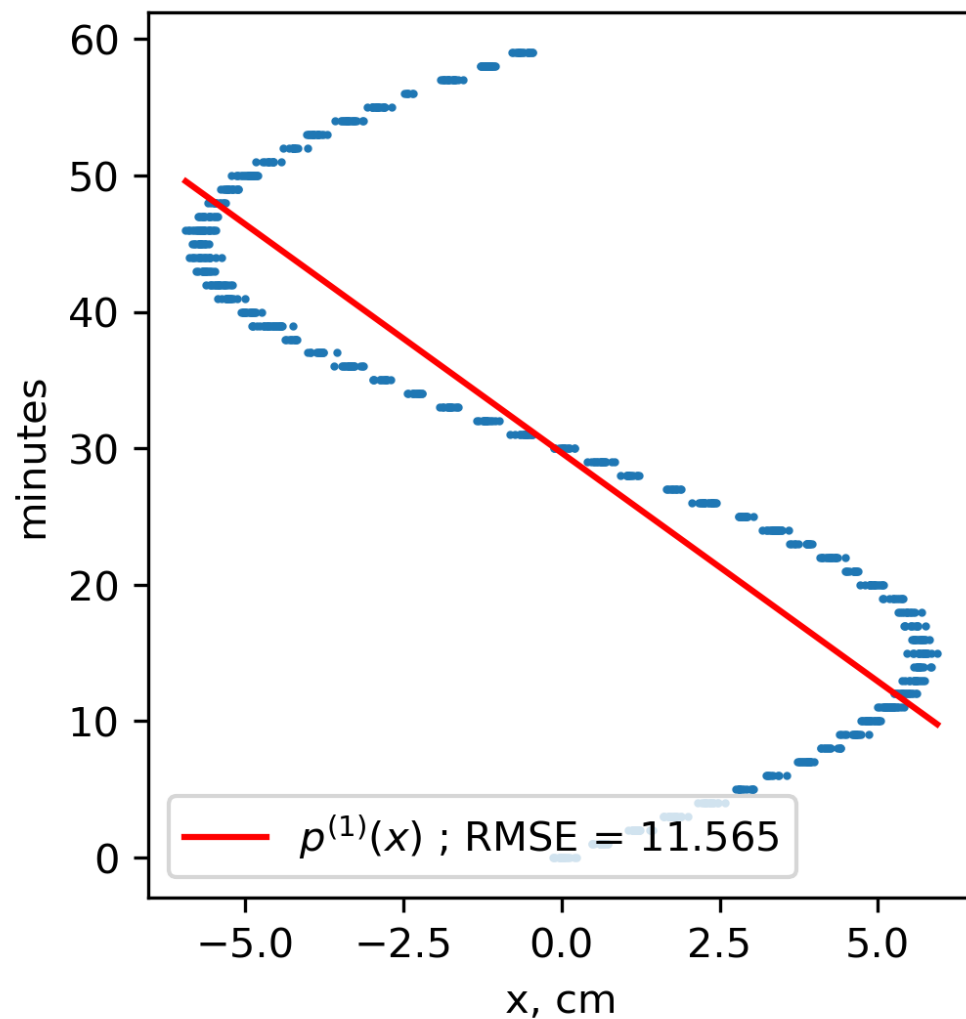
Модель в задаче восстановления регрессии: $\hat{m}_i = f(\vec{p}, x_i) = kx_i + b$

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

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

Решение задачи восстановления регрессии: ПРИМЕР

Результаты модели



Решение задачи восстановления регрессии: ПРИМЕР

Построение и настройка модели

Возьмем модель посильнее

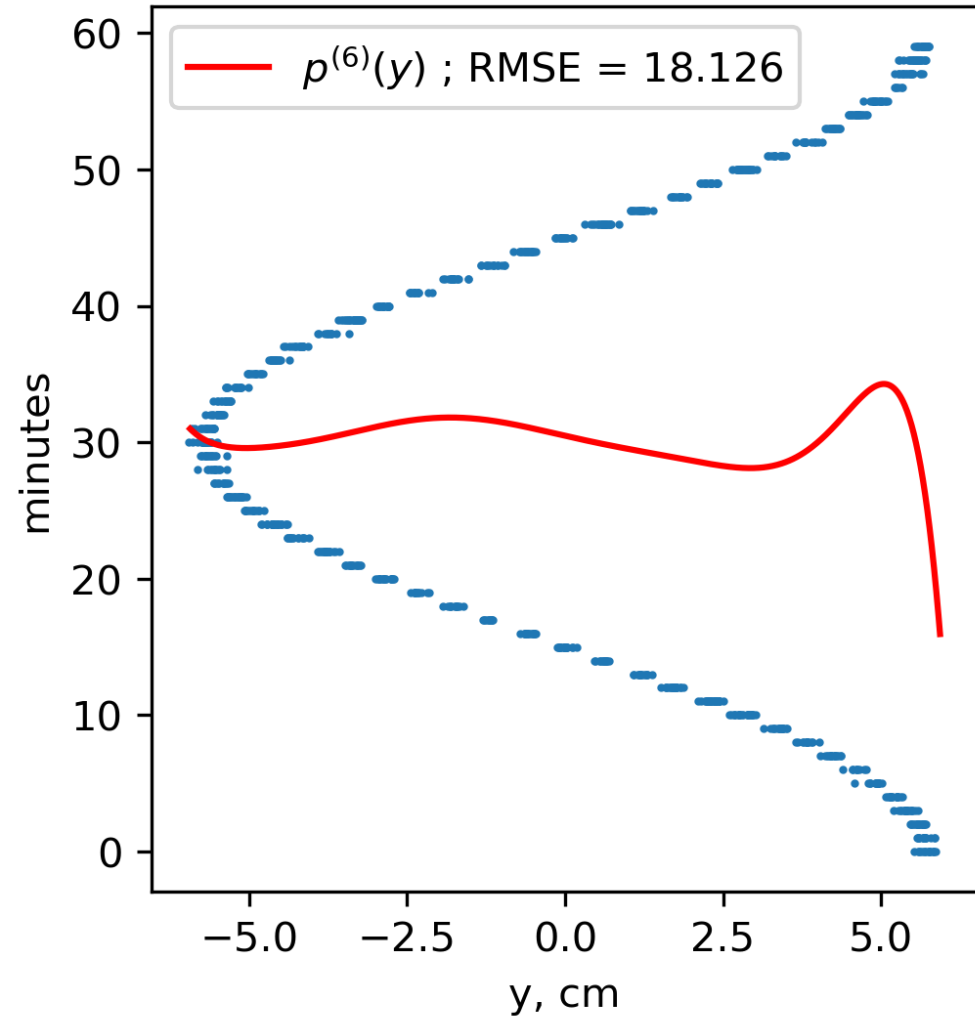
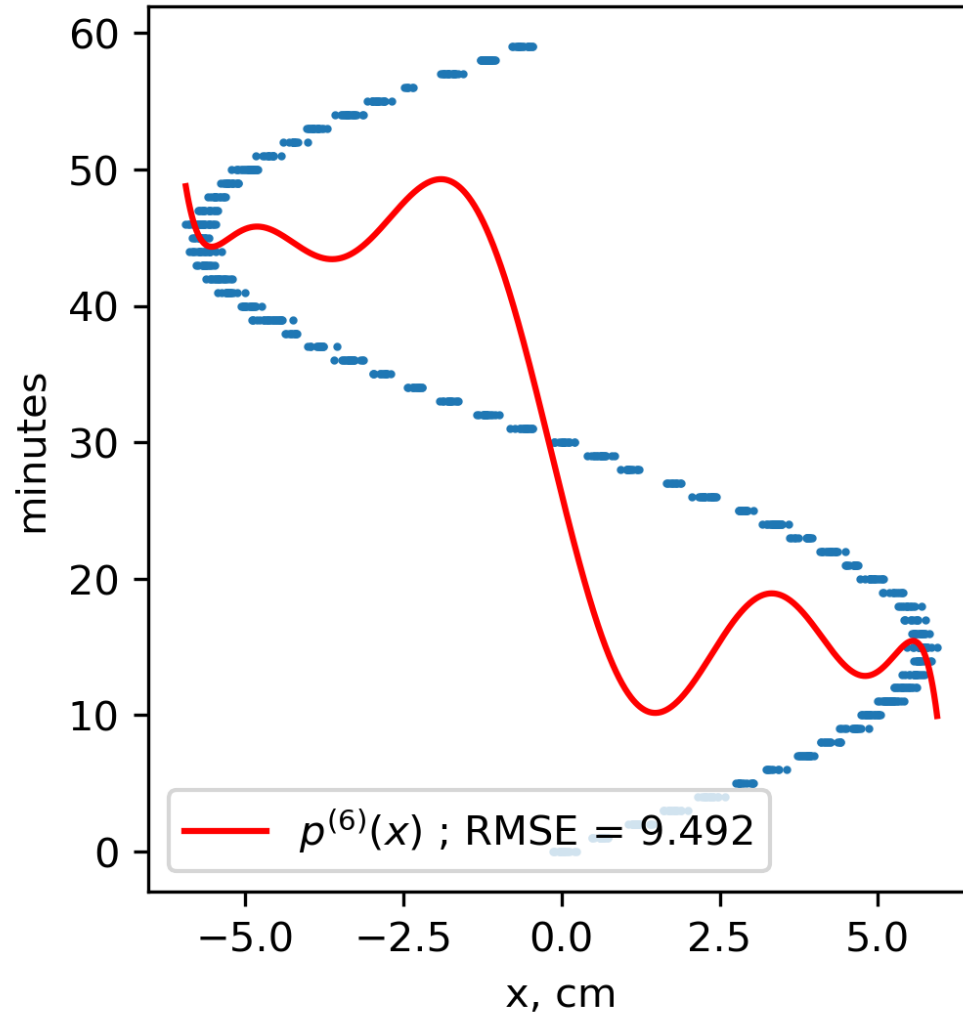
Модель в задаче восстановления регрессии: $\hat{\mathbf{m}}_i = f(\vec{\mathbf{p}}, \mathbf{x}_i) = \text{poly}^{(6)}(\mathbf{x}_i)$

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

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

Решение задачи восстановления регрессии: ПРИМЕР

Результаты модели



Решение задачи восстановления регрессии: ПРИМЕР

Построение и настройка модели

Возьмем нейросеть

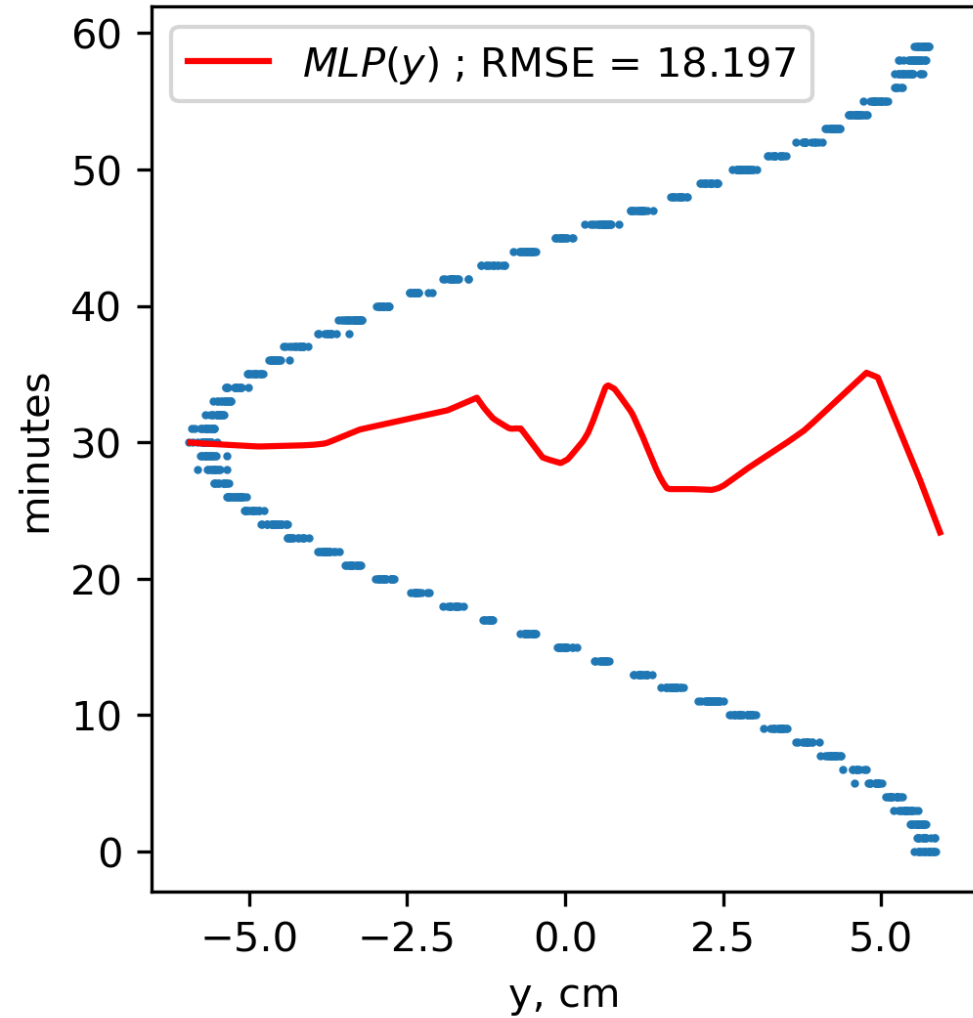
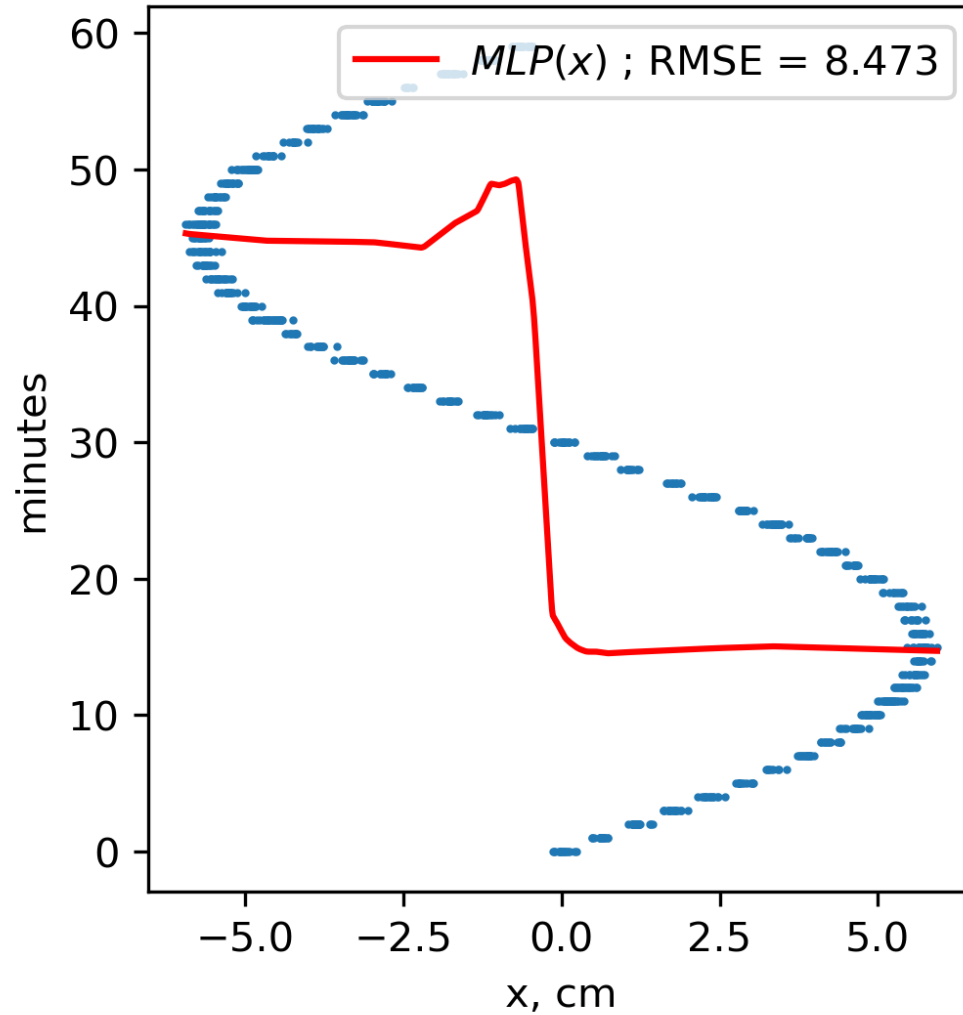
$$\hat{\mathbf{m}}_i = \text{MLP}(\vec{\mathbf{p}}, \mathbf{x}_i)$$

$$\mathcal{L}(\vec{\mathbf{p}}, \{\mathbf{x}_i\}, \{\mathbf{m}_i\}) = \frac{1}{n} \sum_{i=1}^n (\text{MLP}(\vec{\mathbf{p}}, \mathbf{x}_i) - \mathbf{m}_i)^2$$

$$\vec{\mathbf{p}}^* = \underset{\mathbb{P}}{\operatorname{argmin}}(\mathcal{L}(\vec{\mathbf{p}}, \{\mathbf{x}_i\}, \{\mathbf{m}_i\}))$$

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?

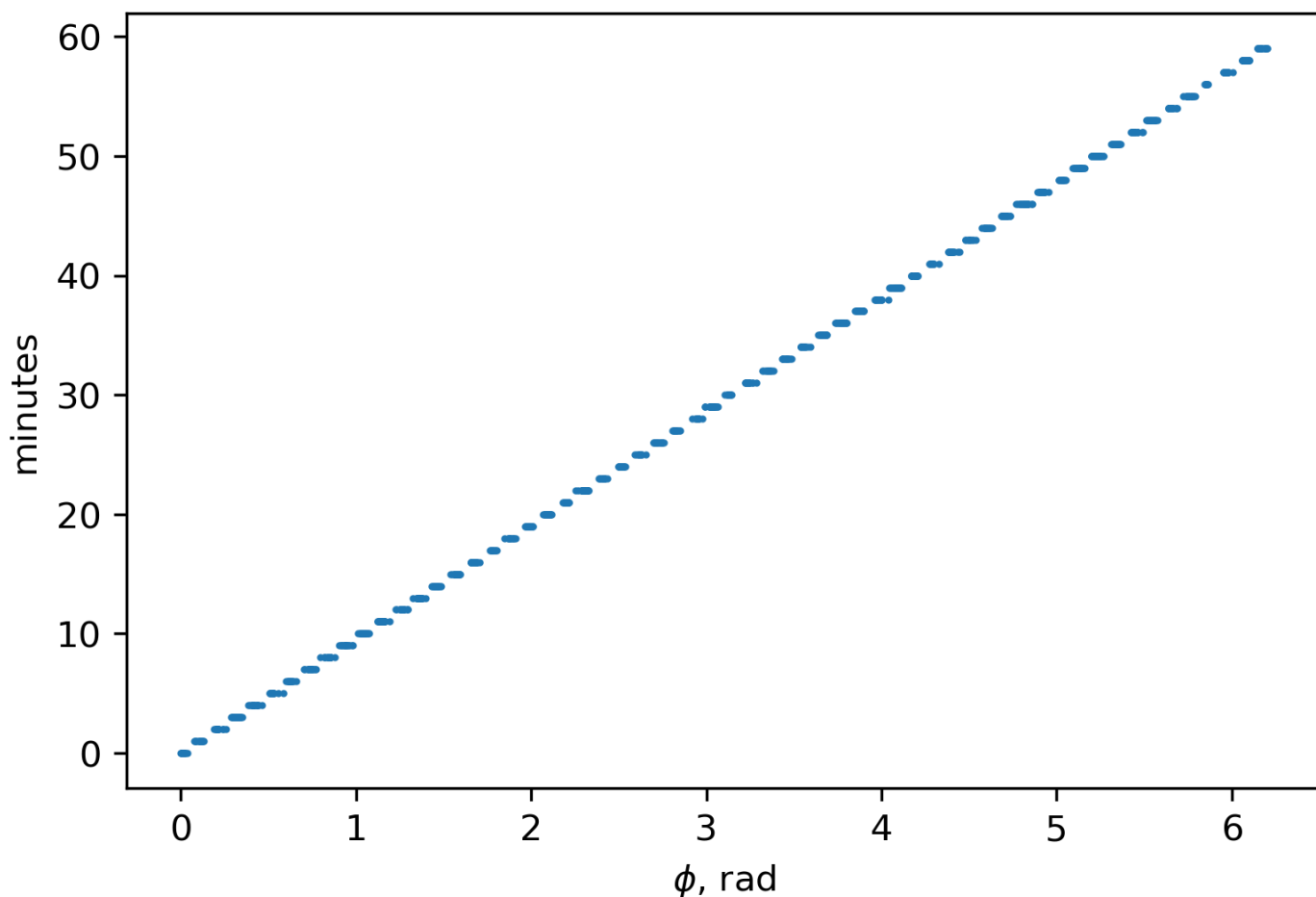
МОЖЕТ, ПРОСТО НЕТ ЗАКОНОМЕРНОСТИ?

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: $\hat{m}_i = f(\vec{p}, \phi_i) = k\phi_i + b$

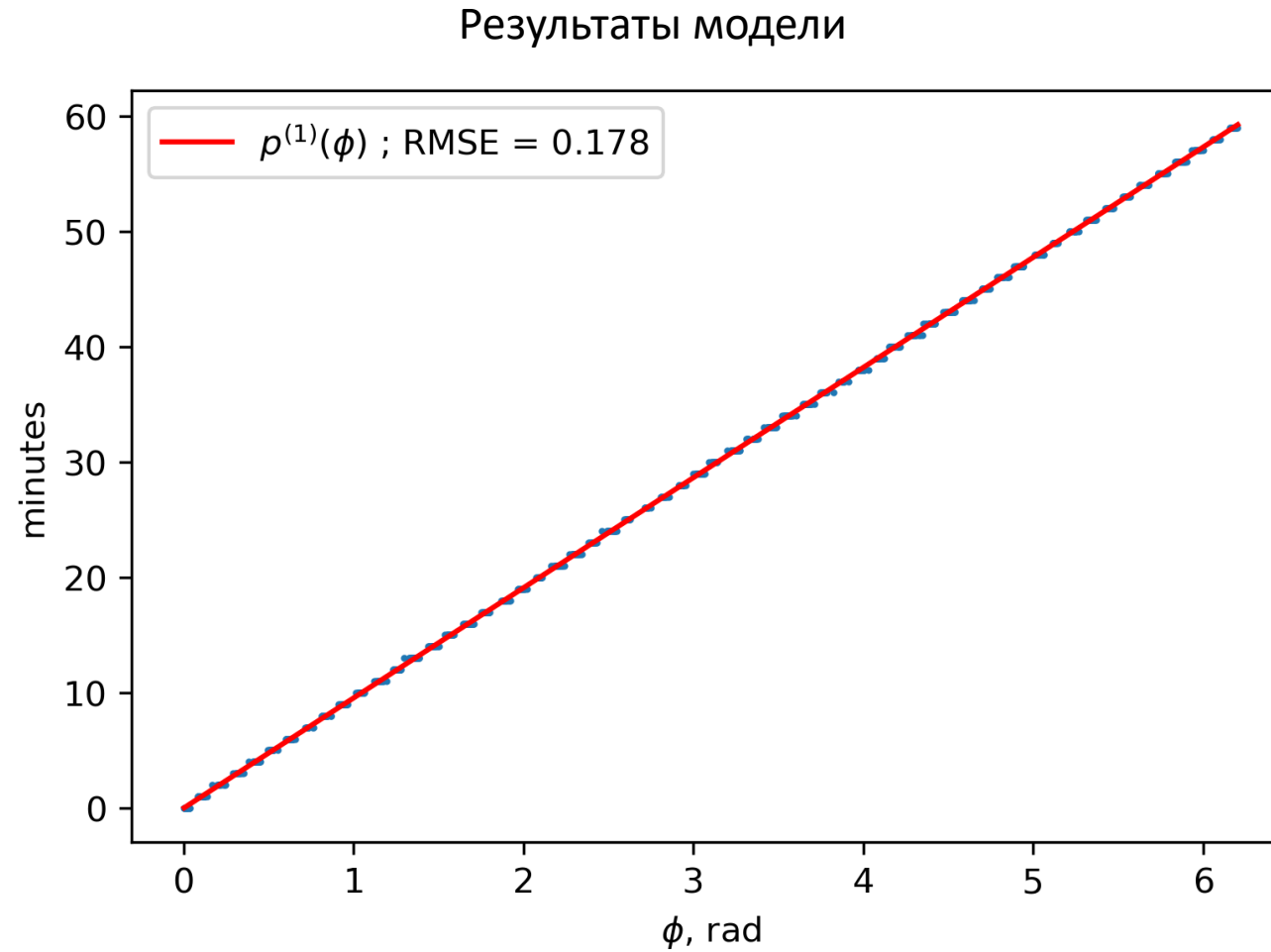
Функция потерь:

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

Решение (оценка параметров \vec{p}):

$$\vec{p}^* = \operatorname{argmin}_{\mathbb{P}} (\mathcal{L}(\vec{p}, \{\phi_i\}, \{m_i\}))$$

Решение задачи восстановления регрессии: ПРИМЕР



Решение задачи восстановления регрессии: ПРИМЕР

использовать (более) полную информацию о событиях

Новое признаковое описание событий: \vec{x}_i - обе координаты x, y конца минутной стрелки

Возьмем нейросеть

$$\hat{m}_i = MLP(\vec{p}, \vec{x}_i)$$

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

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

Качество модели: $RMSE = 0.28m$