



Машинное обучение в науках о Земле

Михаил Криницкий

krinitsky.ma@phystech.edu

К.Т.Н., С.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

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

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

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

один из способов решения:

$\mathcal{L}(\hat{\mathcal{F}}(x))$ — 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

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

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

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

один из способов решения:

$\mathcal{L}(\hat{\mathcal{F}}(x))$ — функционал ошибки

(эмпирического риска, потерь), Loss function

**Чем руководствоваться при выборе
функции ошибки?**

КАКИЕ бывают функции ошибки?!

$\hat{y}_i = \hat{\mathcal{F}}(x_i) = f(\vec{p}, x_i)$ — функционально

задаваемая исследователем, функциональная

исследователя о виде закономерности. Иногда

задается параметрически, \vec{p} — вектор параметров.

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

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

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

Обучение по прецедентам: вероятностная постановка

принцип максимального правдоподобия maximum likelihood estimation

\vec{x}_i - признаковое описание объектов

\vec{y}_i - признаковое описание ответов

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

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

Предположение!

(x_i, y_i) – выбираются из $p(x, y)$
независимо и случайно

Обучение по прецедентам: вероятностная постановка

принцип максимального правдоподобия maximum likelihood estimation

x_i - признаковое описание объектов

y_i - признаковое описание ответов

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

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

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

Предположение!

(x_i, y_i) – выбираются из $p(x, y)$
независимо и случайно

Обучение по прецедентам: вероятностная постановка

принцип максимального правдоподобия maximum likelihood estimation

x_i - признаковое описание объектов
 y_i - признаковое описание ответов
 $p(x, y)$ – (искомая, аппроксимируемая)
совместная плотность распределения
событий на множестве $X \times Y$
 $\phi(x, y, \theta)$ - модель плотности
распределения, предлагаемая
исследователем

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

Предположение!

(x_i, y_i) – выбираются из $p(x, y)$
независимо и случайно

MLE

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

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

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

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

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

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