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

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

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

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

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

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.

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

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

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

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

$\mathcal{T}$: $\{\boldsymbol{x_i}; \boldsymbol{y_i}\}$ — «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)