Введение в машинное обучение

Национальный исследовательский университет "Высшая школа экономики"
Yandex School of Data Analysis

Неофициальный конспект по курсу.

12 апреля 2021 г.

Содержание

1 Метрические методы классификации в задаче восстановления регрессии

1.1 Параметрический подход

Ранее мы рассмотрели *метрические методы классификации* (методы ближайших соседей, окна Парзена и потенциальных функций), основанные на идее измерения *расстояний между объектами*. Эту же идею можно *перенести* и на *задачу восстановления регрессии*.

В задаче регрессии обычно задана обучающая выборка, пара «объект-ответ», в которых ответы — это действительные числа:

- X — объекты (часто $\mathbb{R}^n$); Y — ответы (часто $\mathbb{R}$, реже $\mathbb{R}^m$);
  $X^l = (x_i, y_i)_{i=1}^l$ — обучающая выборка размера l;
  $y_i = y(x_i)$, $y : X \to Y$ — неизвестная зависимость;

И стандартным подходом к решению регрессионных задач является фиксация некоторой параметрической модели зависимости — функция f от объекта x и вектора *параметров* α:

- $a(x) = f(x, \alpha)$ — параметрическая модель зависимости,
  $\alpha \in \mathbb{R}^p$ — вектор параметров модели.

И далее идет процесс *определения вектора параметров* с помощью метода наименьших квадратов. Для этого выписывается *функционал* среднего квадрата ошибки и ставится *оптимизационная задача*: найти вектор параметров, доставляющего этому функционалу минимум:

- Метод наименьших квадратов (МНК):

$$Q(\alpha, X^l) = \sum_{i=1}^{l} w_i \Big(f(x_i, \alpha) - y_i \Big)^2 \to \underset{\alpha}{min},$$

  где w_i — вес, степень важности i-го объекта.

При этом мы в данном случае рассматриваем этот функционал в несколько *обобщенном виде* и ввели веса объектов или *степени важности объектов* обучающей выборки (обычно в функционале такого не делают).

Недостаток:
надо заранее иметь хорошую параметрическую модель $f(x, \alpha)$.

Далеко не всегда в распоряжении исследователей имеются такие модели, поэтому хотелось бы *отходить от параметрического подхода*.

1.2 Непараметрическая регрессия. Формула Надарая-Ватсона

Идеи **непараметрических методов** заключаются в том, чтобы приблизить искомую зависимость константы, но локально в окрестности того объекта x, в котором мы хотим вычислить нашу аппроксимирующую функцию:

Приближение константой $f(x,\alpha)=\alpha$ в окрестности $x\in X$:

$$Q(\alpha;X^l)=\sum_{i=1}^{l} w_i(x)\big(\alpha-y_i\big)^2 \to \min_{\alpha\in\mathbb{R}};$$

где $w_i(x)=K\left(\frac{\rho(x,x_i)}{h}\right)$ — веса объектов x_i относительно x;

$K(r)$ — ядро, невозрастающее, ограниченное, гладкое (желательно);

h — ширина окна сглаживания.

Для этого мы снова пользуемся методом *наименьших квадратов*, а вместо *параметрической модели* зависимости подставляем *константу* α, но теперь вся сложность задачи у нас перемещается в *веса объектов*.

Эти *веса* w_i мы задаем в зависимости от того объекта x, в котором мы ищем значение аппроксимирующей функции. Соответственно, используем *функцию расстояния* ρ (чтобы вес был тем меньше, чем дальше объект x до объекта обучающей выборки x_i), к которой применяем к ней некое *ядро* K. Также мы используем ширину окна h, чтобы варьировать скорость убывания этой функции по мере возрастания расстояния между объектами.

Нам потребуется выразить α. В целом сделать это достаточно просто: у нас функционал $Q(\alpha)$, нам необходимо найти его минимум, продифференцируем по α, приравняем к нулю производную и отсюда найдем α (советуем сделать самостоятельно, как упражнение). Таким образом получаем:

Формула ядерного сглаживания Надарая-Ватсона:

$$a_h(x;X^l)=\frac{\sum_{i=1}^{l} y_i w_i(x)}{\sum_{i=1}^{l} w_i(x)}=\frac{\sum_{i=1}^{l} y_i K\big(\frac{\rho(x,x_i)}{h}\big)}{\sum_{i=1}^{l} K\big(\frac{\rho(x,x_i)}{h}\big)}$$

По сути это просто *средневзвешанное значение ответов* y_i на объектах обучающей выборки. Веса уже зависят от того, *насколько i-й объект далек* от того объекта x, в котором мы вычисляем функцию — *чем дальше, тем меньше вес.*