Семинар 10.
ММП, осень 2012–2013
11 декабря

Илья Толстихин
iliya.tolstikhin@gmail.com

Темы семинара:

- Линейная регрессия;
- Метод наименьших квадратов;
- Геометрическая интерпретация.

1 Линейная регрессия, МНК.

До сих пор мы рассматривали с вами задачи классификации, когда множество ответов $\mathbb{Y}$ было конечным множеством. Теперь мы приступаем к изучению задач классификации, когда $\mathbb{Y} = \mathbb{R}$. Пространство объектов зафиксируем равным $\mathbb{R}^n$.

Мы будем рассматривать *линейные* модели вида

$$a(\boldsymbol{x}) = \alpha_0 + \langle \boldsymbol{x}, \boldsymbol{\alpha} \rangle = \alpha_0 + \sum_{i=1}^{n} \alpha_i x_i.$$

В данном случае линейными модели являются *по параметрам* $(\alpha_0, \boldsymbol{\alpha})$. Часто для простоты мы будем дополнять все объекты обучающей выборки константным признаком, тем самым включив сдвиг α_0 в вектор $\boldsymbol{\alpha}$. Нашей ближайшей задачей будет разобраться, как выбирать параметры модели $\boldsymbol{\alpha}$ по обучающей выборке $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{\ell}$ и к каким геометрическим интерпретациям эти способы ведут.

Как упоминалось на лекции, наиболее часто используемым критерием выбора вектора параметров α является **метод наименьших квадратов**:

$$\sum_{i=1}^{\ell} \bigl(y_i - a_{\boldsymbol{\alpha}}(\boldsymbol{x}_i)\bigr)^2 = \sum_{i=1}^{\ell} (y_i - \langle \boldsymbol{\alpha}, \boldsymbol{x}_i \rangle)^2 \to \min_{\boldsymbol{\alpha}}. \tag{1}$$

Мы можем воспользоваться матричными обозначениями $Y = (y_1, \ldots, y_\ell)^\top$, $A \in \mathbb{R}^{\ell \times n}$ — матрица «объект-признак», тогда задача (1) перепишется в следующем виде:

$$L(\boldsymbol{\alpha}) = \|A\boldsymbol{\alpha} - Y\|^2 = (A\boldsymbol{\alpha} - Y)^\top (A\boldsymbol{\alpha} - Y) \to \min_{\boldsymbol{\alpha}}. \tag{2}$$

Итак, задача заключается в минимизации квадратичного функционала от $\boldsymbol{\alpha}$. Пользуясь элементарными тождествами найдем градиент этого функционала и приравняем его нулю:

$$\frac{\partial L}{\partial \boldsymbol{\alpha}} = 2A^\top A\boldsymbol{\alpha} - 2A^\top Y = 0,$$

откуда получаем условие на решение $\boldsymbol{\alpha}^*$ задачи (1):

$$A^\top A\boldsymbol{\alpha}^* = A^\top Y. \tag{3}$$

В случае, когда матрица $A^\top A$ обратима, мы получим уникальное решение задачи (2):

$$\boldsymbol{\alpha}^* = (A^\top A)^{-1}A^\top Y.$$

Этот алгоритм на обучающей выборке будет выдавать ответы $\hat{Y} = A(A^\top A)^{-1}A^\top Y$.

2 Геометрическая интерпретация МНК.

Оказывается, у решения МНК существует наглядная геометрическая интерпретация. Заметим, что $A\boldsymbol{\alpha} \in \mathbb{R}^\ell$ — линейная комбинация столбцов матрицы A. Обозначим с помощью $\mathrm{Lin}(A)$ линейное подпространство, натянутое на столбцы матрицы A.

Задача. *Найдите $\boldsymbol{\alpha}^*$, для которого $A\boldsymbol{\alpha}^*$ — ортогональная проекция вектора Y на подпространство $\mathrm{Lin}(A)$.*

Мы хотим найти такой вектор $\boldsymbol{\alpha}^*$, для которого вектор $Y - A\boldsymbol{\alpha}^*$ ортогонален всем векторам подпространства $\mathrm{Lin}(A)$. Поскольку произвольный вектор из $\mathrm{Lin}(A)$ выражается в виде $A\boldsymbol{\beta}$ для некоторого $\boldsymbol{\beta} \in \mathbb{R}^n$, то это условие ортогональности легко записывается в следующем виде:

$$\langle Y - A\boldsymbol{\alpha}^*, A\boldsymbol{\beta}\rangle = \boldsymbol{\beta}^\top A^\top (Y - A\boldsymbol{\alpha}^*) = 0, \quad \forall \boldsymbol{\beta},$$

откуда получаем

$$A^\top A\boldsymbol{\alpha}^* = A^\top Y,$$

то есть мы получаем то же условие (3), к которому нас привел МНК. Откуда мы заключаем, что МНК ищет ортогональную проекцию вектора ответов обучающей выборки Y на подпространство столбцов матрицы A.

В каких случаях матрицы $A^\top A$ вырождена? Несложно заметить, что это матрица Грамма системы столбцов матрицы A, поэтому она вырождена тогда и только тогда, когда матрица A имеет неполный столбцовый ранг. В этом случае все геометрические рассуждения остаются справедливыми и МНК действительно найдет проекцию вектора Y на подпространство $\mathrm{Lin}(A)$, но, поскольку столбцы матрицы A линейно зависимы, разложение этой проекции по ним будет не единственным.

Задача. *Рассмотрим задачу с $\mathbb{X} = \mathbb{R}$ и обучающей выборкой из двух объектов: $x_1 = 1$ и $x_2 = 2$ с ответами y_1 и y_2 соответственно. Выпишите решения МНК и посчитайте квадратичную ошибку с и без добавления константного признака.*