Факторизационная машина с локальным низким рангом весовой матрицы

Трофимов Михаил Игоревич

Студент

ФУПМ МФТИ, Москва, Россия

E-mail: `mikhail.trofimov@phystech.edu`

Факторизационные машины [1] представляют собой класс линейных моделей второго порядка, у которых на матрицу весов наложено структурное ограничение – низкий ранг:

$$f(x) = w_0 + \langle \mathbf{w_1}, x \rangle + \sum_{i=0}^{k} \sum_{j=i+1}^{k} W_{ij} x_i x_j$$

$$\mathbf{W} = \mathbf{V}\mathbf{V}^T, V \in R^{d \times k}$$

Подобная модель реализует тот же класс разделяющих поверхностей, что и метод опорных векторов с полиномиальным ядром степени 2. Структурное ограничение на матрицу весов $\mathbf{W}$ можно рассматривать как способ регуляризации, благодаря которому метод оказывается эффективным в задачах большой размерности с разреженным признаковым описанием объектов, в отличие от соответствующей машины опорных векторов.

Однако, недавнее исследование в области матричных аппроксимаций [2] показывает, что гипотеза о низком ранге может быть успешно заменена гиптезой о локальном низком ранге.

Матрицу $\mathbf{W}$ предлагается моделировать как сумму s матриц ранга k:

$$W_{ij} = \sum_{l=1}^{s} p_{ij}^{(l)} W_{ij}^{(l)}, \quad rank(\mathbf{W}^{(l)}) = k, \quad p_{ij}^{(l)} \in [0, 1]$$

Центральная идея заключается в использовании весов $p_{ij}^{(l)}$, которые определяют вклад l-й аппроксимации в точке (i, j). Авторы предлагают использовать поход, аналогичный формуле Надарайя-Ватсона:

$$p_{ij}^{(l)} = \frac{K_h[(i, j), a_l]}{\sum_{t=1}^{s} K_h[(i, j), a_t]}$$

где $K_h[(i, j), (m, n)]$ – функция ядра на индексах элементов матрицы, а $a_q = (i_q, j_q), q \in [1 \ldots l]$ – опорные точки. Данный подход подразумевает много степеней свободы (выбор функции ядра, опорных точек, ранга аппроксимаций и их количества), однако даже эвристический выбор этих параметров позволяет получить результаты, превосходящие классические методы в задаче аппроксимации матрицы.

В данной работе гипотеза о локальном низком ранге применена к весовой матрице в модели факторизационной машины. Адаптируя подход авторов [2], в качестве K_h использовалось ядро Епанечникова:

$$K_h[(i,j),(m,n)] = \frac{3}{4}(1 - d_{(i,j),(m,n)})\mathcal{I}_{[d_{(i,j),(m,n)}<h]}$$

Функция расстояния d была факторизована как произведение расстояний на строках и столбцах, то есть

$$d_{(i,j),(m,n)} = \hat{d}(i,m) \cdot \hat{d}(j,n)$$

$$\hat{d}(i,m) = \arccos(\frac{\langle u_i, u_m \rangle}{||u_i|| \cdot ||u_m||})$$

Представления u_i, используемые в функции расстояния $\hat{d}$, были взяты как соотвествующие векторы из усеченного сингулярного разложения исходной матрицы признаков.

Для настройки полученного метода предлагается использовать вариацию онлайн метода градиентного спуска, поощряющую разреженность коэффициентов решения – FTRL [3]. Полученная модель имеет сложность обучения $O(n)$ относительно n – количества обучающих примеров, что позволяет использовать ее на выборках большого размера.

Эксперименты показывают, что предложенная модель превосходит классическую модель факторизационных машин.

Литература

1. Steffen Rendle Factorization machines with libFM // ACM Trans. Intell. Syst. Technol. May 2012, Vol. 3
2. Joonseok Lee at el Local Low-Rank Matrix Approximation // Proceedings of the 30 th International Conference on Machine Learning, Atlanta, Georgia, USA, 2013. JMLR:W&CP, volume 28
3. H. Brendan McMahan Follow-the-Regularized-Leader and Mirror Descent: Equivalence Theorems and L1 Regularization // Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS) (2011)