InClass Prediction Competition

Movie recomendation TS Spring 2020

Movie recomendation

8 teams · a month ago

Overview

Description
Evaluation

Предсказание рейтинга фильмов

Вам необходимо реализовать алгоритм implicit ALS для обучения факторизационной машины и применить его для предсказания рейтинга фильмов.

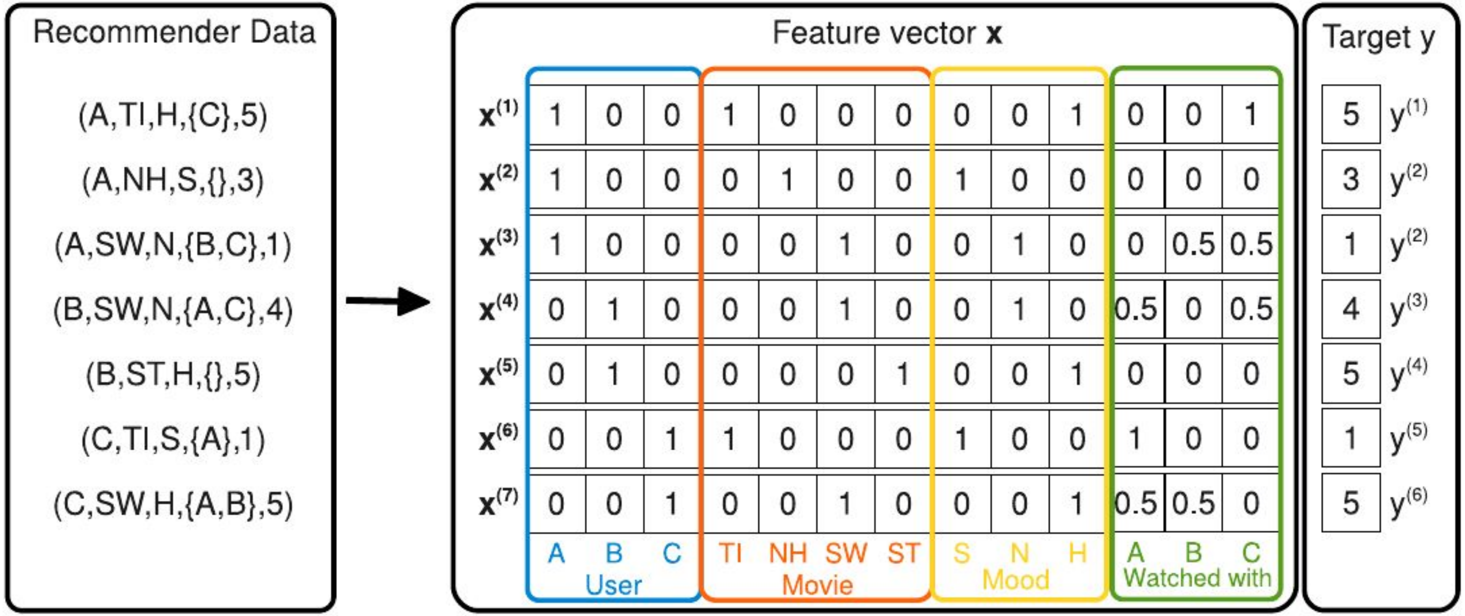
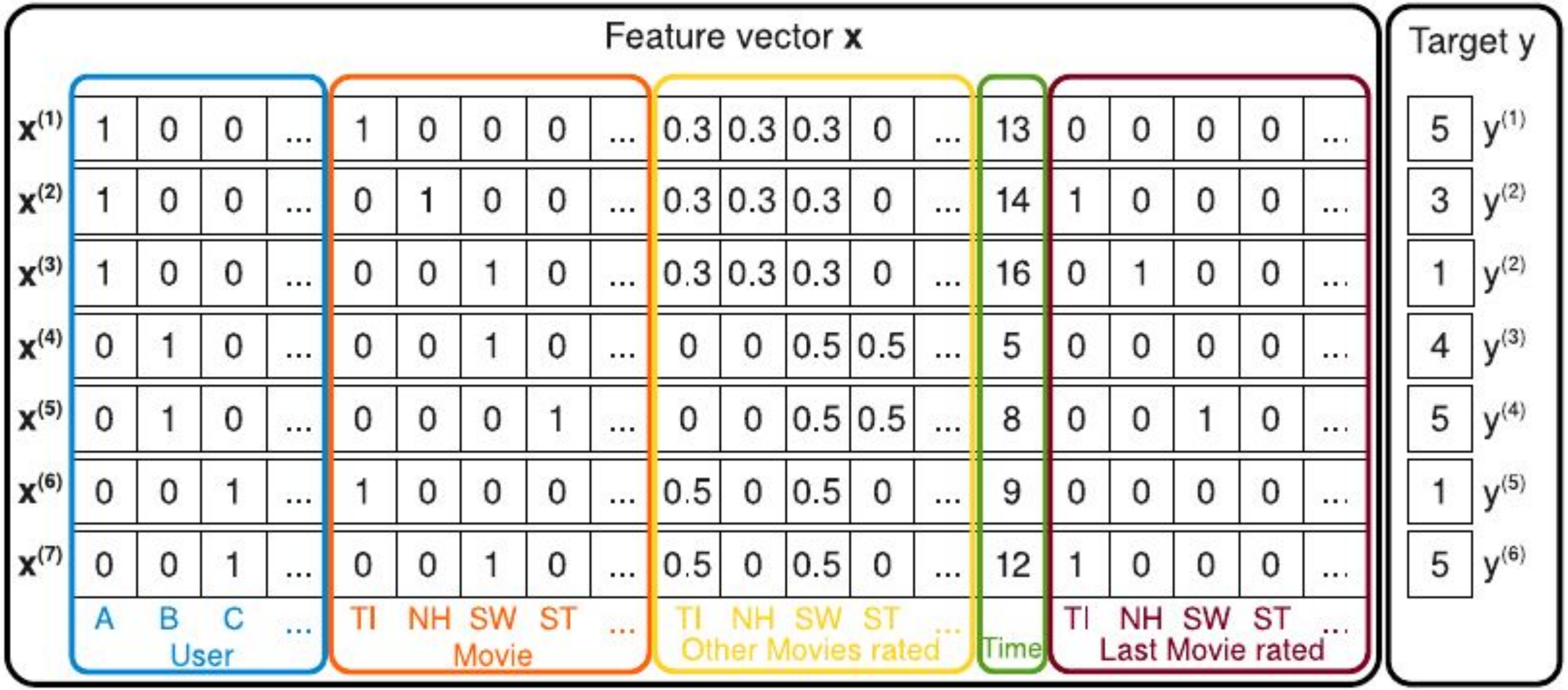


Figure 2: Context-aware recommendation data (left side) is transformed into a prediction problem from real-valued features (right side) by encoding the categorical and set categorical variables (left side) with indicator variables (right side). Here in the feature vector x, the first three values indicate the user, the next four ones the movie, the next three ones the mood and the last three ones the other users a movie has been watched with.

Steffen Rendle, Zeno Gantner, Christoph Freudenthaler, and Lars Schmidt-Thieme. 2011. Fast context-aware recommendations with factorization machines.



Factorization Machines

Steffen Rendle

Department of Reasoning for Intelligence

The Institute of Scientific and Industrial Research

Описание модели

Steffen Rendle, Zeno Gantner, Christoph Freudenthaler, and Lars Schmidt-Thieme. 2011. Fast context-aware recommendations with factorization machines.

1) Model Equation: The model equation for a factorization machine of degree $d = 2$ is defined as:

$$\hat{y}(\mathbf{x}) := w_0 + \sum_{i=1}^{n} w_i \, x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \langle \mathbf{v}_i, \mathbf{v}_j \rangle \, x_i \, x_j \qquad (1)$$

where the model parameters that have to be estimated are:

$$w_0 \in \mathbb{R}, \quad \mathbf{w} \in \mathbb{R}^n, \quad \mathbf{V} \in \mathbb{R}^{n \times k} \qquad (2)$$

And $\langle \cdot, \cdot \rangle$ is the dot product of two vectors of size k:

$$\langle \mathbf{v}_i, \mathbf{v}_j \rangle := \sum_{f=1}^{k} v_{i,f} \cdot v_{j,f} \qquad (3)$$

A row $\mathbf{v}_i$ within $\mathbf{V}$ describes the i-th variable with k factors. $k \in \mathbb{N}_0^+$ is a hyperparameter that defines the dimensionality of the factorization.

A 2-way FM (degree $d = 2$) captures all single and pairwise interactions between variables:

- w_0 is the global bias.
- w_i models the strength of the i-th variable.
- $\hat{w}_{i,j} := \langle \mathbf{v}_i, \mathbf{v}_j \rangle$ models the interaction between the i-th and j-th variable. Instead of using an own model parameter $w_{i,j} \in \mathbb{R}$ for each interaction, the FM models the interaction by factorizing it. We will see later on, that this is the key point which allows high quality parameter estimates of higher-order interactions ($d \geq 2$) under sparsity.

ALS алгоритм

Steffen Rendle, Zeno Gantner, Christoph Freudenthaler, and Lars Schmidt-Thieme. 2011. Fast context-aware recommendations with factorization machines.

$$\begin{aligned} h_{(v_{l,f})}(\mathbf{x}) &= x_l \sum_{i=1}^{n} v_{i,f}\, x_i - x_l^2\, v_{l,f} \\ &= x_l\, q(\mathbf{x}, f|\Theta) - x_l^2\, v_{l,f} && (14) \\ & && (15) \end{aligned}$$

with

$$q(\mathbf{x}, f|\Theta) := \sum_{i=1}^{n} v_{i,f}\, x_i \qquad (16)$$

$$e(x, y|\Theta^*) = e(x, y|\Theta) + (\theta^* - \theta)\, h_{(\theta)}(\mathbf{x}) \qquad (13)$$

procedure LEARNALS(S)
- $w_0 \leftarrow 0$ ▷ Initialize the model parameters
- $\mathbf{w} \leftarrow (0, \ldots, 0)$
- $\mathbf{V} \sim \mathcal{N}(0, \sigma)$
- **for** $(\mathbf{x}, y) \in S$ **do** ▷ Precompute e and q
  - $e(\mathbf{x}, y|\Theta) \leftarrow \hat{y}(\mathbf{x}, y) - y$
  - **for** $f \in \{1, \ldots, k\}$ **do**
    - $q(\mathbf{x}, f|\Theta) \leftarrow \sum_{i=1}^{n} v_{i,f}\, x_i$
  - **end for**
- **end for**
- **repeat** ▷ Main optimization loop
  - $w_0^* \leftarrow -\frac{\sum_{(\mathbf{x},y)\in S} (e(\mathbf{x},y|\Theta) - w_0)}{|S| + \lambda_{(w_0)}}$ ▷ global bias
  - $e(\mathbf{x}, y|\Theta^*) \leftarrow e(\mathbf{x}, y|\Theta) + (w_0^* - w_0)$
  - $w_0 \leftarrow w_0^*$
  - **for** $l \in \{1, \ldots, n\}$ **do** ▷ 1-way interactions
    - $w_l \leftarrow -\frac{\sum_{(\mathbf{x},y)\in S} (e(\mathbf{x},y|\Theta) - w_l\, x_l)\, x_l}{\sum_{(\mathbf{x},y)\in S} x_l^2 + \lambda_{(w_l)}}$
    - $e(\mathbf{x}, y|\Theta^*) \leftarrow e(\mathbf{x}, y|\Theta) + (w_l^* - w_l)\, x_l$
    - $w_l \leftarrow w_l^*$
  - **end for**
  - **for** $f \in \{1, \ldots, k\}$ **do** ▷ 2-way interactions
    - **for** $l \in \{1, \ldots, n\}$ **do**
      - $v_{l,f}^* \leftarrow -\frac{\sum_{(\mathbf{x},y)\in S} \left(e(\mathbf{x},y|\Theta) - v_{l,f}\, h_{(v_{l,f})}(\mathbf{x})\right) h_{(v_{l,f})}(\mathbf{x})}{\sum_{(\mathbf{x},y)\in S} h^2_{(v_{l,f})}(\mathbf{x}) + \lambda_{(v_{l,f})}}$
      - $e(\mathbf{x}, y|\Theta^*) \leftarrow e(\mathbf{x}, y|\Theta) + (v_{l,f}^* - v_{l,f})\, x_l$
      - $q(\mathbf{x}, f|\Theta^*) \leftarrow q(\mathbf{x}, f|\Theta) + (v_{l,f}^* - v_{l,f})\, x_l$
      - $v_{l,f} \leftarrow v_{l,f}^*$
    - **end for**
  - **end for**
- **until** stopping criterion is met
- **return** $w_0, \mathbf{w}, \mathbf{V}$

end procedure

Figure 3: Alternating least algorithm that optimizes the model parameters w_0, $\mathbf{w}$ and $\mathbf{V}$ for least-square in $O(|S|\,\overline{m}_{|S|}\,k)$ time (see section 4.3.3) where $|S|$ are the number of training examples and $\overline{m}_{|S|}$ the average number of non-zero elements in an input vector $\mathbf{x}$.

Параметры обучения

n_iter	16
latent dimension	5
lambda w	4
lambda v	7

Инициализация весов

w_0 - средняя оценка фильма по всей базе

v_0 - [средняя оценка пользователя, средняя оценка фильма,

средняя оценка пользователя, средняя оценка фильма]

V - Xavier Glorot initialization

$$W \sim U\left[-\frac{\sqrt{6}}{\sqrt{n_j + n_{j+1}}}, \frac{\sqrt{6}}{\sqrt{n_j + n_{j+1}}}\right] \qquad (16)$$

$w_0 \rightarrow w_0 / 5$ $\qquad$ $v_0 \rightarrow v_0 / 5$

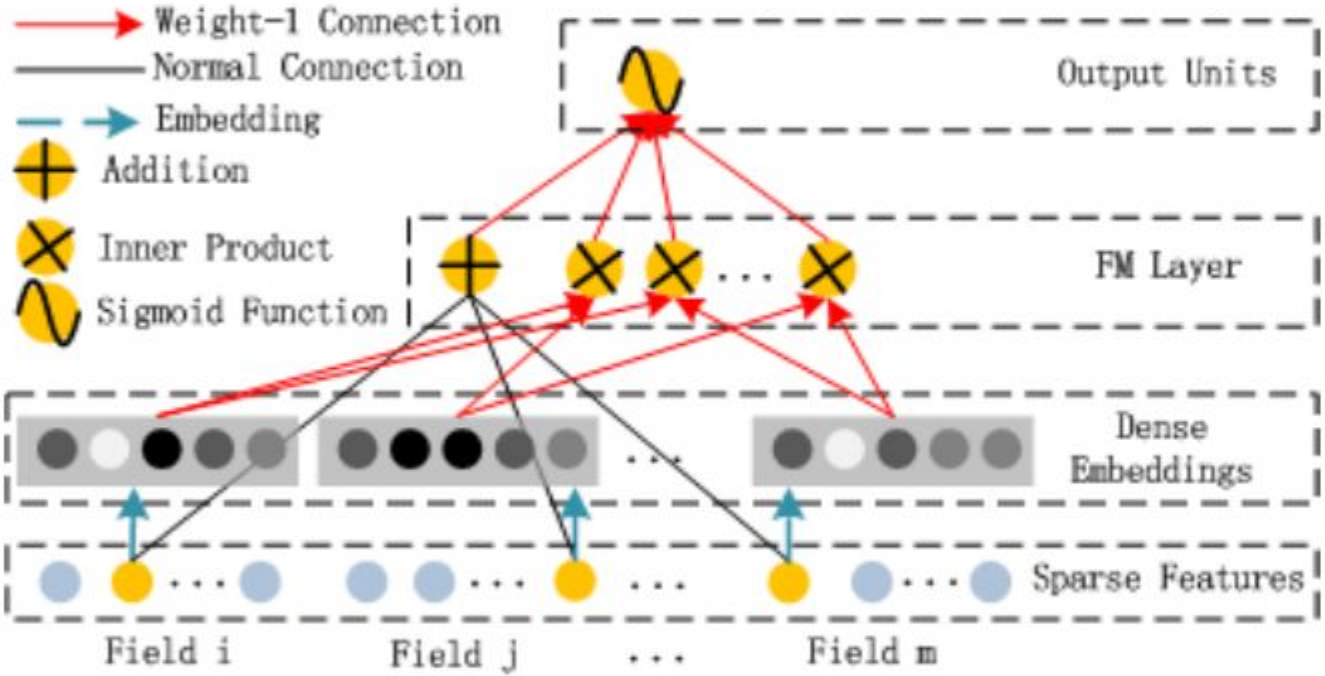


Figure 2: The architecture of FM.