Complex Networks and Statistical Learning
Homework 2

Chenghua Liu
liuch18@mails.tsinghua.edu.cn
Department of Computer Science
Tsinghua University

目录

1 Markov Network

Consider the following scenario: in movie rating websites, movies tend to get ratings that reflect whether each user likes the movie or not. For simplicity, let us assume that ratings are binary(like or dislike). We design the following Markov Network to capture the behavior of users in a movie rating website:

(a) Each movie is associated with two variables: category and general reception, both arediscrete variables, there are K categories ofmovies and general reception variable has 5values (bad, normal, good, very good,exceptional)

(b) Each user is associated with one variable,the user archetype. There are G archetypesfor the users

(c) The rating distribution for each (user, movie) pair depends on all three variables(movie reception, movie category and userarchetype)

(d) Movie category variables are observed,ratings are observed, some movies havetheir general reception labeled initially, allother variables are hidden

(e) There are a total of N users and M movies

Choose proper notations for the variables, andspecify the formal Markov Network formulation.Derive update formulas for the inference of hidden variables, and based on the inference procedure describe how to learn the unknown parameters of the Markov Network
solution:

1.1 Defination

We set random variables as follows. Rating given by user i to movie j $R_{ij} \in$ {-1 if dislike, +1 if like} . Archetype of user i $A_i \in \{1, \ldots, G\}$. Category and General reception of movie j $C_j \in \{1, \ldots, K\}, G_j \in \{1, \ldots, 5\}$.

We define notation as follows. $\mathcal{S} = \{(i, j)\}$: Set of (i, j) where user i has rated movie j or where we want to estimate the rating given by user i to movie j; $\mathcal{B}(\cdot)$: the Markov blanket of a given set of random variables.We mark lowercase letters denote the values of the correspondig random variables and bold letters denote the collection of the corresponding variables.

For a rating r given by a user of archetype a to a movie of category c and general reception g, the factor is:

$$f(r, a, c, g) := r\theta_{a,c,g} = \begin{cases} -\theta_{a,c,g}, & \text{if } r = -1 \\ +\theta_{a,c,g}, & \text{if } r = +1 \end{cases}$$

where $\theta_{a,c,g}$'s are parameters.And the joint-distribution is :

$$P(\boldsymbol{r}, \boldsymbol{a}, \boldsymbol{c}, \boldsymbol{g}) = \frac{1}{Z} \mathrm{e}^{E(\boldsymbol{r}, \boldsymbol{a}, \boldsymbol{c}, \boldsymbol{g})} = \frac{1}{Z} \exp\left(\sum_{(i,j)\in\mathcal{S}} f\left(r_{ij}, a_i, c_j, g_j\right)\right) = \frac{1}{Z} \exp\left(\sum_{(i,j)\in\mathcal{S}} r_{ij}\theta_{a_i,c_j,g_j}\right)$$

1.2 Derivation

Now,let's derive update formulas for the inference of hidden variables.According to gibbs sampling, we have inference for user archetypes:

$$P\left(a_i \mid \mathcal{B}\left(a_i\right)\right)=\frac{\exp \left(\sum_{j:(i,j) \in \mathcal{S}} f\left(r_{ij}, a_i, c_j, g_j\right)\right)}{\sum_{x=1}^G \exp \left(\sum_{j:(i,j) \in \mathcal{S}} f\left(r_{ij}, x, c_j, g_j\right)\right)}=\frac{\exp \left(\sum_{j:(i,j) \in \mathcal{S}} r_{ij} \theta_{a_i, c_j, g_j}\right)}{\sum_{x=1}^G \exp \left(\sum_{j:(i,j) \in \mathcal{S}} r_{ij} \theta_{x, c_j, g_j}\right)}$$

Inference for general reception:

$$P\left(g_j \mid \mathcal{B}\left(g_j\right)\right)=\frac{\exp \left(\sum_{i:(i,j) \in \mathcal{S}} f\left(r_{ij}, a_i, c_j, g_j\right)\right)}{\sum_{z=1}^5 \exp \left(\sum_{i:(i,j) \in \mathcal{S}} f\left(r_{ij}, a_i, c_j, z\right)\right)}=\frac{\exp \left(\sum_{i:(i,j) \in \mathcal{S}} r_{ij} \theta_{a_i, c_j, g_j}\right)}{\sum_{z=1}^5 \exp \left(\sum_{i:(i,j) \in \mathcal{S}} r_{ij} \theta_{a_i, c_j, z}\right)}$$

and inference for ratings:

$$P\left(r_{ij} \mid \mathcal{B}\left(r_{ij}\right)\right)=\frac{\exp \left(f\left(r_{ij}, a_i, c_j, g_j\right)\right)}{\sum_{s \in\{-1,+1\}} \exp \left(f\left(s, a_i, c_j, g_j\right)\right)}=\frac{\exp \left(r_{ij} \theta_{a_i, c_j, g_j}\right)}{\sum_{s \in\{-1,+1\}} \exp \left(s \theta_{a_i, c_j, g_j}\right)}$$

When it is difficult to sample directly from the distribution, we could take variational inference method. We assume that the mean-field

$$Q(\boldsymbol{r}, \boldsymbol{a}, \boldsymbol{c}, \boldsymbol{g})=\left(\prod_{i=1}^N Q_{A_i}\left(a_i\right)\right)\left(\prod_{j=1}^M Q_{C_j}\left(c_j\right) Q_{G_j}\left(g_j\right)\right)\left(\prod_{(i,j) \in \mathcal{S}} Q_{R_{ij}}\left(r_{ij}\right)\right)$$

Then we have inference for user archetypes:

$$Q_{A_i}\left(a_i\right) \propto \exp \left(\sum_{j:(i,j) \in \mathcal{S}} \mathbb{E}_{Q_{R_{ij}, C_j, G_j}} f\left(R_{ij}, a_i, C_j, G_j\right)\right)=\exp \left(\sum_{j:(i,j) \in \mathcal{S}} \mathbb{E}_{Q_{R_{ij}}}\left[R_{ij}\right] \cdot \mathbb{E}_{Q_{C_j} Q_{G_j}}\left[\theta_{a_i, C_j, G_j}\right]\right)$$

Inference for movie reception:

$$Q_{G_j}\left(g_j\right) \propto \exp \left(\sum_{i:(i,j) \in \mathcal{S}} \mathbb{E}_{Q_{R_{ij}, A_i, C_j}} f\left(R_{ij}, A_i, C_j, g_j\right)\right)=\exp \left(\sum_{i:(i,j) \in \mathcal{S}} \mathbb{E}_{Q_{R_{ij}}}\left[R_{ij}\right] \cdot \mathbb{E}_{Q_{A_i} Q_{C_j}}\left[\theta_{A_i, C_j, g_j}\right]\right)$$

And inference for ratings:

$$Q_{R_{ij}}\left(r_{ij}\right) \propto \exp \left(\mathbb{E}_{Q_{A_i, C_j, G_j}} f\left(r_{ij}, A_i, C_j, G_j\right)\right)=\exp \left(r_{ij} \cdot \mathbb{E}_{Q_{A_i}} Q_{C_j} Q_{G_j}\left[\theta_{A_i, C_j, G_j}\right]\right)$$

1.3 Learning

We need to estimate the distributions of hidden variables by method mentioned above before parameter learning.After estimation of hidden variables,consider the log-likelihood $\log P(\boldsymbol{\theta})=\sum_{(i,j) \in \mathcal{S}} f\left(r_{ij}, a_i, c_j, g_j\right)-\log Z$. For the first term, the gradient is:

$$\frac{\partial f\left(r_{ij}, a_i, c_j, g_j\right)}{\partial \theta_{a, c, g}}=r_{ij} \cdot \frac{\partial \theta_{a_i, c_j, g_j}}{\partial \theta_{a, c, g}}=r_{ij} \cdot 1_{\left\{a_i=a, c_j=c, g_j=g\right\}}$$

For the second term $\log Z$, the gradient is

$$\frac{\partial \log Z}{\partial \theta_{a,c,g}} = \frac{1}{Z} \sum_{\boldsymbol{r},\boldsymbol{a},\boldsymbol{c},\boldsymbol{g}} \mathrm{e}^{E(\boldsymbol{r},\boldsymbol{a},\boldsymbol{c},\boldsymbol{g})} \frac{\partial E(\boldsymbol{r},\boldsymbol{a},\boldsymbol{c},\boldsymbol{g})}{\partial \theta_{a,c,g}} = \mathbb{E}_P \left[\frac{\partial E(\boldsymbol{R},\boldsymbol{A},\boldsymbol{C},\boldsymbol{G})}{\partial \theta_{a,c,g}} \right] = \sum_{(i,j)\in\mathcal{S}} \mathbb{E}_P \left[\frac{\partial f\left(R_{ij}, A_i, C_j, G_j\right)}{\partial \theta_{a,c,g}} \right]$$

$$= \sum_{(i,j)\in\mathcal{S}} \mathbb{E}_P \left[R_{ij} \cdot 1_{\{A_i=a, C_j=c, G_j=g\}} \right]$$

Then we can use stochastic gradient descent to parameter learning.

2 Markov Logic Network

Consider again the scenario in the last problem,this time we attempt to use Markov Logic Networkto solve it.

(a) Specify the schema of the relational dataset(what are the relations, tuples, domains,attributes in the dataset)

(b) Use First-Order Logic to capture thedistributional information as in the lastproblem

(c) If we use conditional likelihood maximization to learn weights, which set ofvariables should be treated as context variables?

(d) Is there any opportunity to use lifted inference in this particular problem setting?

solution:

2.1 Schema

2.1.1 Relations

We set ”Users” as the data of each user;”Movies” as the data of each movie; and ”Ratings” as the data of each rating given by a user to a movie.

2.1.2 Attributes

In relation Users,we set ”UserID” (primary key) as the ID i of the user; ”Archetype” as the archetype a_i of user i.

In relation Movies,we set ”MovieID” (primary key) as the ID j of the movie; ”Category” as the category c_j of movie j; ”Reception” as the general reception g_j of movie j.

In relation Ratings,we set ” UserID” (foreign key) as the user i who gave the rating; ”MovieID” (foreign key) as the movie j to which the rating is given; ”Rating” as the rating r_{ij} given by user i to movie j.

2.1.3 Tuples

In relation Users: (UserID, Archetype) = (i, a_i), the data of user i.

In relation Movies: (MovieID, Category, Reception) = (j, c_j, g_j), the data of movie j.

In relation Ratings: (UserID, MovieID, Rating) = (i, j, r_{ij}), the data of the rating given by user i to movie j.

2.1.4 Domains

In relation Users,we set UserID: integer in $\{1, \ldots, N\}$; Archetype: integer in $\{1, \ldots, G\}$.

In relation Movies,we set MovieID: integer in $\{1, \ldots, M\}$; Category: integer in $\{1, \ldots, K\}$; Reception: integer in $\{1, \ldots, 5\}$.

In relation Ratings,we set UserID: integer in $\{1, \ldots, N\}$; MovieID: integer in $\{1, \ldots, M\}$; Rating: integer in $\{-1, +1\}$.

2.2 First-Order Logic

Let $\mathbf{P} = \{(r, a, c, g)\}$ be the set of movie preference information, i.e., a user of archetype a prefers to give rating r to a movie of category c and general reception g. The set $\mathbf{P}$ could be constructed based on prior knowledge or experience. For each $(r, a, c, g) \in \mathbf{P}$, we can construct a first order logic expression as:

(Ratings.UserID = Users.UserID) (Ratings.MovieID = Movies.MovieID) (Users.Archetype =a) (Movies.Category = c) (Movies.Reception=g) ⇒ (Ratings.Rating =r).

2.3 Context Variables

We should treat movie categories as context variables because that only movie categories are always observed.

2.4 Lifted Inference

Using lifted inference in this particular problem setting will be more efficient. Because the distribution of a rating r_{ij} only depends on a_i, c_j and g_j for each possible (a, c, g) tuple, which means that we can inference the ratings r_{ij} at the same time for all (i, j) pairs. Therefore, the number of possible (a, c, g) tuples should be much less than the number of missing ratings.