



Grupo de Pesquisa em Reconhecimento
de Padrões e Otimização



DISCIPLINA: PGENE 613 - Tópicos Avançados em Aprendizado de Máquina e Otimização

Tópicos em Aprendizado de Máquina e Otimização

2020-2

Trabalho 1: Autoencoders e Sistema de Recomendação

Deseja-se projetar um sistema de recomendação para usuários de filmes utilizando a base de dados MovieLens na sua versão *small*, que tem as seguintes características, conforme o site que disponibiliza a mesma:

“This dataset (ml-latest-small) describes 5-star rating and free-text tagging activity from [MovieLens](<http://movielens.org>), a movie recommendation service. It contains 100836 ratings and 3683 tag applications across 9742 movies. These data were created by 610 users between March 29, 1996 and September 24, 2018. This dataset was generated on September 26, 2018.”

No drive estão disponibilizados os arquivos dessa base:

movings.csv – contém a lista de filmes

ratings.csv – contém as notas atribuídas pelos usuários

readmet.txt – informações da base

Vale observar que, embora sejam atribuídas notas a 9742 filmes, a numeração dos filmes na tabela ratings vai até 193609.

No final sugere-se o pseudocódigo de um algoritmo para implementação desse sistema de recomendação, em conformidade com a teoria abordada em sala de aula.

Pede-se que:

- 1) Os valores dos parâmetros taxa de aprendizado α e fator de regularização λ devem ser propostos pelo aluno. Recomenda-se apenas que o valor para o fator de regularização seja da ordem de 10 vezes menor do que a taxa de aprendizado. Pede-se citar os valores que foram utilizados.
- 2) Pede-se utilizar uma arquitetura de autoencoder com apenas uma camada de neurônios escondido. O número k de neurônios escondido deve ser uma escolha do aluno. Sugere-se simular 3 valores de k , numa faixa entre 50 e 100.
- 3) Encerrar a simulação quando um valor do erro médio quadrático MSE for menor do que 0,1.

DISCIPLINA: PGENE 613 - Tópicos Avançados em Aprendizado de Máquina e Otimização

- 4) Pede-se que seja elaborada uma tabela com 8 linhas e 6 colunas. Nas 8 linhas da tabela devem constar o número dos seguintes indivíduos: 40, 92, 123, 245, 312, 460, 514 e 590. Nas 5 primeiras colunas de cada indivíduo devem constar os nomes de 5 filmes sugeridos para os mesmos (filmes com maiores notas para o indivíduo). Os filmes sugeridos não podem incluir aqueles para os quais o usuário já atribuiu uma nota. **Os filmes sugeridos têm que estar na lista dos 9742 filmes aos quais foram atribuídas notas.** Pede-se ainda que, numa sexta coluna, para cada indivíduo, coloque-se o erro médio quadrático das previsões feitas pelo sistema para os filmes que o mesmo atribuiu nota.

Estrutura sugerida para o Relatório:

1. Introdução
2. Revisão bibliográfica (comente sobre um artigo que utilize autoencoders
3. Fundamentação teórica (explique o que é um autoencoder e a técnica de treinamento de autoencoders empilhados)
4. Resultados
5. Conclusões
6. Referências bibliográficas

Algorithm

begin

Inicializar randomicamente as matrizes W e V

$S = \{(i,j); (entrada\ i \neq 0, saída\ j \neq 0)\};$

while not convergence **do**

begin

Misturar randomicamente as entradas de S

for each $(i,j) \in S$ *in shuffled order* **do**

begin

$$e_{ij} = y_j - \sum_{q=1}^K w_{qi} v_{jq}$$

for each $q \in \{1, \dots, k\}$ **do** $w_{qi}^+ = w_{qi} \cdot (1 - 2\alpha\lambda) + 2\alpha e_{ij} \cdot v_{jq}$

for each $q \in \{1 \dots k\}$ **do** $v_{jq}^+ = v_{jq} \cdot (1 - 2\alpha\lambda) + 2\alpha e_{ij} \cdot w_{qi}$

for each $q \in \{1 \dots k\}$ **do** $u_{qi} = u_{qi}^+$ and $v_{jq} = v_{jq}^+$

end

check convergence condition

end

end

$R^T = V \cdot W$

Em que: k = n. de neurônios na camada intermediária.