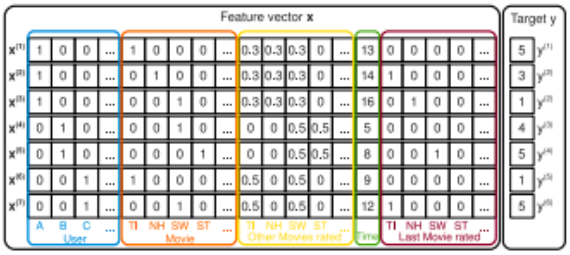
Product Recommendation Engines

Business Case
Financial firms offer a wide array of products: checking accounts and retirement plans are just the beginning. Today, you have joined the ranks of the data scientists at Santander. Your current website provides a large number of recommendations to a small number of customers, which means that your company is not taking full advantage of the data sets your currently own and maintain. If you could recommend the right product to the right customer, you could accelerate your growth and meet the real demands of your consumer base.

Machine Learning Method
In this project, you will learn about Factorization Machines. This is a highly scalable algorithm that was developed by Steffen Rendel in 2010. It has the capacity to leverage extremely large datasets at the Terabyte scale, while still training in linear time.

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

In essence, the factorization machines model is calculating the dot product between the user information and the item information, then computing the difference between those to update the model.



In order to accomplish this, you'll need to **format your data as events**. Each row in your final data set will need to be a single point in time when a customer interacts with a product account. Each column will be either a binary indicator for the product/user, or another feature.

Data set description
At your disposal, you have 1.5 years of consumer behavior. Your data begins in January 2015, and includes monthly records of customer activity. This includes credit card ownerships, savings account utilization etc.

https://www.kaggle.com/c/santander-product-recommendation/data

Your goal is to prediction additional products a customer will purchase in the final month, June 2016, given what they already have in the previous month. These products are the columns named: ind_(xyz)_ult1, which are the columns #25 - #48 in the training data.

The test and train sets are split by time.

	Column Name	Description
1	fecha_dato	The table is partitioned for this column
2	ncodpers	Customer code
3	ind_empleado	Employee index: A active, B ex employed, F filial, N not employee, P pasive
4	pais_residencia	Customer's Country residence
5	sexo	Customer's sex
6	age	Age
7	fecha_alta	The date in which the customer became as the first holder of a contract in the bank
8	ind_nuevo	New customer Index. 1 if the customer registered in the last 6 months.
9	antiguedad	Customer seniority (in months)
10	indrel	1 (First/Primary), 99 (Primary customer during the month but not at the end of the month)
11	ult_fec_cli_1t	Last date as primary customer (if he isn't at the end of the month)
12	indrel_1mes	Customer type at the beginning of the month ,1 (First/Primary customer), 2 (co-owner),P (Potential),3 (former primary), 4(former co-owner)
13	tiprel_1mes	Customer relation type at the beginning of the month, A (active), I (inactive), P (former customer),R (Potential)
14	indresi	Residence index (S (Yes) or N (No) if the residence country is the same than the bank country)
15	indext	Foreigner index (S (Yes) or N (No) if the customer's birth country is different than the bank country)
16	conyuemp	Spouse index. 1 if the customer is spouse of an employee
17	canal_entrada	channel used by the customer to join
18	indfall	Deceased index. N/S
19	tipodom	Addres type. 1, primary address
20	cod_prov	Province code (customer's address)
21	nomprov	Province name
22	ind_actividad_cliente	Activity index (1, active customer; 0, inactive customer)
23	renta	Gross income of the household
24	segmento	segmentation: 01 - VIP, 02 - Individuals 03 - college graduated
25	ind_ahor_fin_ult1	Saving Account
26	ind_aval_fin_ult1	Guarantees
27	ind_cco_fin_ult1	Current Accounts
28	ind_cder_fin_ult1	Derivada Account
29	ind_cno_fin_ult1	Payroll Account
30	ind_ctju_fin_ult1	Junior Account
31	ind_ctma_fin_ult1	Más particular Account
32	ind_ctop_fin_ult1	particular Account
33	ind_ctpp_fin_ult1	particular Plus Account
34	ind_deco_fin_ult1	Short-term deposits
35	ind_deme_fin_ult1	Medium-term deposits
36	ind_dela_fin_ult1	Long-term deposits
37	ind_ecue_fin_ult1	e-account
38	ind_fond_fin_ult1	Funds
39	ind_hip_fin_ult1	Mortgage
40	ind_plan_fin_ult1	Pensions
41	ind_pres_fin_ult1	Loans
42	ind_reca_fin_ult1	Taxes
43	ind_tjcr_fin_ult1	Credit Card
44	ind_valo_fin_ult1	Securities

45	ind_viv_fin_ult1	Home Account
46	ind_nomina_ult1	Payroll
47	ind_nom_pens_ult1	Pensions
48	ind_recibo_ult1	Direct Debit

Code snippets

References

- Intuitively: https://www.analyticsvidhya.com/blog/2018/01/factorization-machines/
- Formally: https://www.csie.ntu.edu.tw/~b97053/paper/Rendle2010FM.pdf