

Пользователь имеет ежедневные данные о продажах. Задача состоит в том, чтобы спрогнозировать общее количество товаров, проданных в каждом магазине. Список магазинов и товаров немного меняется каждый месяц. Данные записаны в файлы формата .csv, который открывается в текстовых редакторах и в Microsoft Office Excel для более наглядного представления данных. Файлы следующие:

- 1) sales_train.csv – набор для обучения. Ежедневные исторические данные с января 2013 года по октябрь 2015 года;
- 2) test.csv – набор тестов. Необходимо спрогнозировать продажи этих магазинов и товаров на ноябрь 2015 года;
- 3) items.csv – дополнительная информация о товарах/продуктах;
- 4) item_categories.csv – дополнительная информация о категориях товаров;
- 5) shops.csv – дополнительная информация о магазинах.

Данные представлены следующим образом:

item_name	item_id	item_category_id
!ВО ВЛАСТИ НАВАЖДЕНИЯ (ПЛАСТ.) D	0	40
"!АВВУУ FineReader 12 Professional Edition Full [PC, Цифровая версия]"	1	76
.....		
Яйцо дракона (Игра престолов)	22169	69

Таблица 1. Данные файла items.csv

item_category_name	item_category_id
PC - Гарнитуры/Наушники	0
Аксессуары - PS2	1
Аксессуары - PS3	2
.....	
Элементы питания	83

Таблица 2. Данные файла item_categories.csv

date	date_block_num	shop_id	item_id	item_price	item_cnt_day
02.01.2013	0	59	22154	999.0	1.0
03.01.2013	0	25	2552	899.0	1.0
04.01.2013	0	25	2572	239.0	1.0
.....					
31.10.2015	33	25	7409	299.0	1.0

Таблица 3. Данные файла sales_train.csv

shop_name	shop_id
"!Якутск Орджоникидзе, 56 фран"	0
"!Якутск ТЦ ""Центральный"" фран"	1
"Адыгея ТЦ ""Мега"""	2
"Балашиха ТРК ""Октябрь-Киномир"""	3
.....	
"Ярославль ТЦ ""Альтаир"""	59

Таблица 4. Данные файла shops.csv

У самого датасета следующие поля с данными:

- 1) ID - идентификатор, представляющий кортеж (Магазин, товар) в тестовом наборе;
- 2) shop_id - уникальный идентификатор магазина;
- 3) item_id - уникальный идентификатор товара;
- 4) item_category_id - уникальный идентификатор категории товара;
- 5) item_cnt_day - количество проданных товаров.
Именно это необходимо спрогнозировать, ежемесячную сумму этого поля;
- 6) item_price - текущая цена товара;
- 7) date - дата в формате дд/мм/гггг;
- 8) date_block_num - номер месяца, используемый для удобства. Январь 2013 года равен 0, февраль 2013 года равен 1. Следовательно, октябрь 2015 года равен 33 номеру;
- 9) item_name - название товара;
- 10) shop_name - название магазина;
- 11) item_category_name - название категории товара.

У обработанной таблицы, готовой для последующего прогнозирования, должны быть следующие поля:

- 1) ID – идентификатор номера в таблице;
- 2) shop_id – уникальный идентификатор магазина;
- 3) shop_name – название магазина;
- 4) item_id – уникальный идентификатор товара;
- 5) item_name – название товара;
- 6) month-year – количество проданных товаров в данном магазине
(данные должны выводиться по месяцам, то есть: 01-2013, 02-2013 и т.д.)

date	shop_name	\
0	Химки ТЦ "Мега"	
1	Цифровой склад 1С-Онлайн	
2	Химки ТЦ "Мега"	
3	Химки ТЦ "Мега"	
4	Химки ТЦ "Мега"	

date	item_name	2013-01	2013-02	\
0	! ВО ВЛАСТИ НАВАЖДЕНИЯ (ПЛАСТ.)	D	0	0
1	!АВВУУ FineReader 12 Professional Edition Full...		0	0
2	***В ЛУЧАХ СЛАВЫ (UNV)	D	0	0
3	***ГОЛУБАЯ ВОЛНА (Univ)	D	0	0
4	***КОРОБКА (СТЕКЛО)	D	0	0

date	2013-03	2013-04	2013-05	2013-06	2013-07	2013-08	2013-09	2013-10	\
0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0

date	2013-11	2013-12	2014-01	2014-02	2014-03	2014-04	2014-05	2014-06	\
0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	2	0	0	0
2	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0

date	2014-07	2014-08	2014-09	2014-10	2014-11	2014-12	2015-01	2015-02	\
0	0	0	1	0	0	0	0	0	0
1	1	1	1	1	0	0	0	0	0
2	0	1	0	0	1	0	0	0	0
3	1	1	0	0	0	0	0	0	0
4	0	0	1	0	0	0	0	0	0

date	2015-03	2015-04	2015-05	2015-06	2015-07	2015-08	2015-09	2015-10
0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0

Как примерно должна выглядеть таблица для дальнейшего прогнозирования

Сам датасет необходимо разделить на тренировочную и тестовые выборки в соотношении 2:1 (первые два года взять для обучения, оставшиеся 12 месяцев – для тестирования). После обучения на тренировочной выборке вывести график ошибки обучения, а после вывода графика продолжить прогнозирование, прибавляя к датасету по месяцу

месяц1 -> месяц2

месяц1 + месяц2 -> месяц3

месяц1 + месяц2 +месяц3 -> месяц4 и т.д.

После получения финального прогноза вывести его в таблице со следующими столбцами:

ID – shop_name – item_name – item_cnt_month

(саму таблицу необходимо записать в .csv файл)

При создании нейросети реализовать произвольное добавление слоёв (поскольку сеть должна быть многослойной, то реализовать посредством ввода числа слоёв внутри программы (например, number_of_layers = число), а также задаваемое пользователем внутри программы количества эпох для обучения (number_of_epochs = число)

После построения прогноза вывести график точности предсказаний (x – месяц, y – суммарные продажи за месяц) и график ошибок и потерь во время обучения модели