Predict Value Amount, by Charge and Retail Features, Oct. 2015

Import GraphLab Create

In [1]:

```
import graphlab
```

Load SFrame

In [2]:

```
itv = graphlab.SFrame('itv.csv')
```

```
[INFO] This non-commercial license of GraphLab Create is assigned to crucker@mediacomcc.comand will expire o
n October 22, 2016. For commercial licensing options, visit https://dato.com/buy/.

[INFO] Start server at: ipc:///tmp/graphlab_server-2893 - Server binary: /usr/local/lib/python2.7/dist-packa
ges/graphlab/unity_server - Server log: /tmp/graphlab_server_1446220056.log
[INFO] GraphLab Server Version: 1.6.1
PROGRESS: Finished parsing file /home/ubuntu/coursera-notebooks/itv.csv
PROGRESS: Parsing completed. Parsed 100 lines in 0.19469 secs.
------------------------------------------------------
Inferred types from first line of file as
column_type_hints=[int,int,int,int]
If parsing fails due to incorrect types, you can correct
the inferred type list above and pass it to read_csv in
the column_type_hints argument
------------------------------------------------------
PROGRESS: Finished parsing file /home/ubuntu/coursera-notebooks/itv.csv
PROGRESS: Parsing completed. Parsed 250000 lines in 0.172378 secs.
```

Show SFrame

In [3]:

```
itv
```

Out[3]:

SUB_ACCT_NO_ITV	MONTHLY_CHRG_AMT_ITV	MONTHLY_RETAIL_AMT_ITV	MONTHLY_VALUE_AMT_ITV
8383100010002051	4	4	4
8383100010002051	2	2	2
8383100010003968	31	38	31
8383100010003968	1	2	1
8383100010003968	4	4	4
8383100010003968	7	7	7
8383100010003968	4	4	4
8383100010003968	0	0	0
8383100010003968	3	3	3
8383100010003968	8	8	8

[250000 rows x 4 columns]
Note: Only the head of the SFrame is printed.
You can use print_rows(num_rows=m, num_columns=n) to print more rows and columns.

Train Model

In [4]:

```
train_data,test_data = itv.random_split(.8,seed=0)
```

Build Model

In [5]:

```
reg_model = graphlab.linear_regression.create(train_data,target='MONTHLY_VALUE_AMT_ITV',features=['MONTHLY_
CHRG_AMT_ITV','MONTHLY_RETAIL_AMT_ITV'])
```

```
PROGRESS: Creating a validation set from 5 percent of training data. This may take a while.
          You can set ``validation_set=None`` to disable validation tracking.

PROGRESS: Linear regression:
PROGRESS: --------------------------------------------------------
PROGRESS: Number of examples          : 190139
PROGRESS: Number of features          : 2
PROGRESS: Number of unpacked features : 2
PROGRESS: Number of coefficients    : 3
PROGRESS: Starting Newton Method
PROGRESS: --------------------------------------------------------
PROGRESS: +-----------+----------+--------------+--------------------+----------------------+---------------
+-----------------+
PROGRESS: | Iteration | Passes   | Elapsed Time | Training-max_error | Validation-max_error | Training-rmse
| Validation-rmse |
PROGRESS: +-----------+----------+--------------+--------------------+----------------------+---------------
+-----------------+
PROGRESS: | 1         | 2        | 1.033084     | 381.280049         | 247.799303           | 3.720650
| 4.017707        |
PROGRESS: +-----------+----------+--------------+--------------------+----------------------+---------------
+-----------------+
```

Evaluate Model

In [6]:

```
print reg_model.evaluate(test_data)
```

```
{'max_error': 247.7993025191349, 'rmse': 3.654877580057717}
```

In [7]:

```
reg_model.get('coefficients')
```

Out[7]:

name	index	value
(intercept)	None	-0.0935131643027
MONTHLY_CHRG_AMT_ITV	None	0.9373111121
MONTHLY_RETAIL_AMT_ITV	None	0.0161227943745

[3 rows x 3 columns]

Explore Account

In [27]:

```
account = itv[itv['SUB_ACCT_NO_ITV']==8383100010092813]
```

In [28]:

```
account
```

Out[28]:

SUB_ACCT_NO_ITV	MONTHLY_CHRG_AMT_ITV	MONTHLY_RETAIL_AMT_ITV	MONTHLY_VALUE_AMT_ITV
8383100010092813	2	2	0

[? rows x 4 columns]
Note: Only the head of the SFrame is printed. This SFrame is lazily evaluated.
You can use len(sf) to force materialization.

In [29]:

```
print account['MONTHLY_VALUE_AMT_ITV']
```

```
[0, ... ]
```

Apply Model

In [30]:

```
reg_model.predict(account)
```

Out[30]:

```
dtype: float
Rows: 1
[1.8133546486467926]
```