Assignment 7

In assignment 5 you will use association rules and clustering to assess fraudulent behavior in simulated Medicare data.

Download the assignment data from Dropbox:
https://www.dropbox.com/s/t9kz2xeai6bw7in/assignment_5_data.zip?dl=0

Turn in a single word document to Blackboard with your **brief but accurate** answers.

1.) Find fraudulent behavior patterns using association rules.

The transaction data set contains the medical procedures of a sample of patients from the larger, general population. While it may contain some fraudulent behavior, it presumably contains what could be called normal patient behavior.

The transaction_review data set contains the medical procedures of 5000 patients who have been manually labeled as involved in Medicare fraud.

The proc_key_map data set contains the English names of medical procedures. Please report results for question number one using these English procedure names.

(2 pts.) Using the SAS Enterprise Miner Association node on all defaults, create a link graph of normal patient behavior using the transaction data set and briefly describe the highest lift rules.

(2 pts.) Using the SAS Enterprise Miner Association node, set minimum confidence level = 5% and support percentage = 1%. Create a link graph of fraudulent patient behavior using the transaction_review data set and briefly describe the highest lift rules.

Specifically use global_proc_id as target and id as id. (Unless you are 100% sure you joined the proc_key_map data set onto the both transaction data sets correctly – if you did not do this join, or you did not do it correctly, your answers will be wrong – if you are 100% sure you did this join correctly, you may use procedure_code as the target.)

Either way please make sure that at least the top few rules have some kind of label in the link graph. You may add the labels as a table outside the plot itself.

(1 pt.) How could these findings be used to detect fraud in the future?

2.) Profile fraudulent patient populations using clustering.

The patient_history data set contains basic demographic information about the normal population sample and SVD features extracted from their procedures.

The patient_history_review data set contains basic demographic information about the fraudulent patient population and SVD features extracted from their procedures.

Since patient procedures can be used to detect fraud, build clusters in the SVD features already created from the transaction sets. Create 13 clusters in the patient_history set using the SAS Enterprise Miner Cluster node. Use only SVD_feature1, SVD_Feature2, and SVD_feature3 as clustering inputs. Aside from

Assignment 7

specifying 13 clusters, set standardization=none and seed initialization method=full replacement. Separately create 6 clusters in the patient_history_review set using the SAS Enterprise Miner Cluster node. Use only SVD_feature1, SVD_Feature2, and SVD_feature3 as clustering inputs. Aside from specifying 6 clusters, set standardization=none and seed initialization method= full replacement.

(1 pt.) Explore the exported data from the cluster node. To confirm the clustering results in the patient_history data set, create a 3-D plot using SVD_feature1 as X, SVD_feature2 as Y, and SVD_feature3 as Z, and _SEGMENT_ as a color variable.

(1 pt.) Explore the exported data from the cluster node. To confirm the clustering results in the patient_history_review data set, create a 3-D plot using SVD_feature1 as X, SVD_feature2 as Y, and SVD_feature3 as Z, and _SEGMENT_ as a color variable.

Use the following SAS code in a SAS Code node directly after the Cluster node(s).

```
proc freq
      data=&EM_IMPORT_DATA;
      table _SEGMENT_*age;
      table _SEGMENT_*gender;
run;
```

*(1 pt.) What can you say about the difference in **age** profiles between the patient_history and patient_history_review data sets?*

*(1 pt.) What can you say about the difference in **gender** profiles between the patient_history and patient_history_review data sets?*

(1 pt.) How could these findings be used to detect fraud in the future?

(FYI: My diagram looked like this when I was finished.)

Assignment 7