Financial (Stock) Data Collection and Storage

By Steve Champion

Contents

- Current Status
- Pros of New Database
- Cursory Analysis
- Future Plans
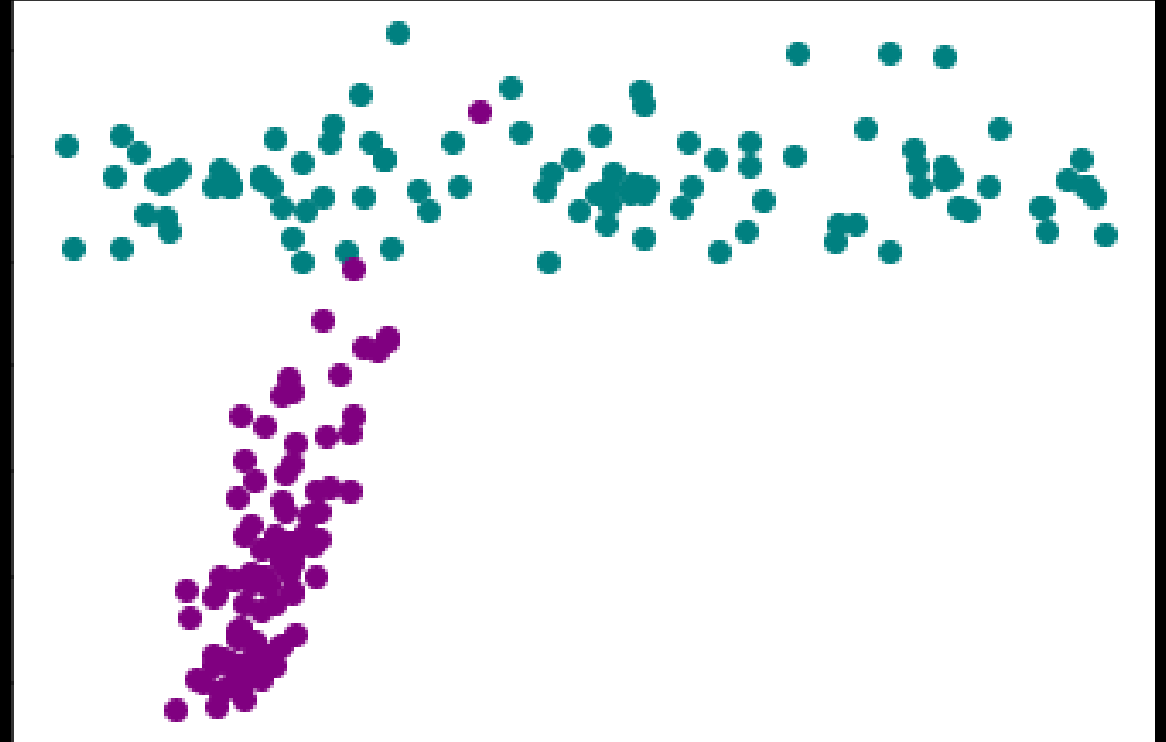- Conclusions

Current Stock Data

Name	Data Format/Acquirement Method	Currentness	Historic Data Granularity	Historic Options Data	Cost
Bloomberg Terminal		Real Time			$24,000/yr
Quandl	API Interface				Variable
Yahoo Finance	Raw Data	Up to 20 minute delay	Day	No	Free
Google Finance	Raw Data	Up to 20 minute delay	Day	No	Free
ToS TD Ameritrade	API Interface	Real Time		No	Free with Account

The Database

- Data comes from Yahoo Finance
- Stored with data science in mind
    - Cleaned
    - Program for simple feature adds
    - Programs for data adds
- Collects Historic Options Data
    - Collects data daily after market close

Cursory Results

- Simple GAN does not perform very well
  - Generated data is not close to real data
  - No predictive ability
  - Needs work

Future Plans

- Develop sentiment analysis programs
  - Create features from sentiment data and add into database
- Develop TGAN
  - Mature GAN theory and update model
  - Multi-featured
- Build alternate models

Conclusion

- Stock data has many sources and formats
  - Easily organized and attainable data will make future model building much easier
- GAN models are complicated
  - Multi-layered Deep Learning Models
  - Much theory behind them

Email any questions to
stevenchamp11@yahoo.com