INTERACTIVE APPLICATIONS IN JUPYTER NOTEBOOKS

Chakri Cherukuri
Romain Menegaux

Quantitative Financial Research Group
OUTLINE

● Introduction to widget ecosystem in Jupyter Notebook

● Applications:
  − Brain imaging (Neuroscience)
  − Twitter Sentiment Analysis (Machine Learning)
  − Yield Curve PCA (Statistics/Finance)
  − Data analysis of U.S. ETFs (Data Analysis/Finance)

● Resources
WIDGET MODEL

Python object

- Interface the user sees
- Its attributes (traits) can send events when changed

Visual Representation

- Graphics, interaction
- JavaScript implementation that is hidden from the user
MAGNETOENCEPHALOGRAPHY (MEG)
EXPERIMENT

Present subject with images, then record brain activity

- Data from the open dataset SPM faces
- Process it using open-source package MNE
- Original talk: Scipy 2017 MNE-Python to See the Brain at a Millisecond Time-Scale https://github.com/agramfort/mne_scipy2017
Each sensor yields a time series

(Spatial coloring)
BRAIN ACTIVATIONS: 2D

Browse through the time series using the IndexSelector
BRAIN ACTIVATIONS: 3D

Using the 3d-plotting library pythreeJS

MEG signal

Magnetic Field (nT)

Time (ms)
TWITTER SENTIMENT MODEL PERFORMANCE ANALYSIS
TWITTER SENTIMENT

Problem statement:

Predict the sentiment (negative, neutral, positive) of a tweet for a company

Ex: “$CTIC Rated strong buy by three WS analysts. Increased target rom $5 to $8.”: Positive

- Can be used as a trading signal
  - Buy stocks with positive sentiment
  - Sell stocks with negative sentiment
CLASSIFICATION PROBLEM

● Input: raw tweets

● Output: sentiment label \( \sum \{ \text{negative, neutral, positive} \} \)

● Methodology
  - Train classifier on training data
  - Use the trained model to predict labels on test data
  - Evaluate performance on the test data
ONE-VS-REST LOGISTIC REGRESSION

- Train three binary classifiers for each label
  - Model 1: Negative vs. Not Negative
  - Model 2: Neutral vs. Not Neutral
  - Model 3: Positive vs. Not Positive

- Get probabilities (measures of confidence) for each label

- Output the label with the maximum probability
CLASSIFIER PERFORMANCE ANALYSIS

- Look at misclassified examples
  - Confusion matrix

- Understand model predicted probabilities
  - Triangle visualization

- Fix data issues
CONFUSION MATRIX

- K x K matrix where K = number of classes
- Cell[i, j] = number of samples whose:
  - Actual label = i
  - Predicted label = j
- Diagonal entries - correct predictions
- Off diagonal entries - misclassifications
TRIANGLE VISUALIZATION

- Model returns 3 probabilities (which sum to 1)
- How can we visualize these 3d “points”?
- Points inside an equilateral triangle
## DASHBOARD COMPONENTS

<table>
<thead>
<tr>
<th>Component</th>
<th>Plotting Widget</th>
<th>Action</th>
<th>Interactivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confusion Matrix</td>
<td>Grid Heat Map (bqplot)</td>
<td>Click on each cell to select tweets in that cell</td>
<td>Data Grid and Triangle will be updated accordingly with selected tweets</td>
</tr>
<tr>
<td>Raw Tweets</td>
<td>Data Grid (custom)</td>
<td>Click on a row to select the specific tweet</td>
<td>Pie and Bar chart will be updated with probabilities and model weights respectively</td>
</tr>
<tr>
<td>Predicted Probabilities</td>
<td>Triangle (bqplot)</td>
<td>Lasso Selector</td>
<td>Data Grid will be updated with lasso selected tweets</td>
</tr>
<tr>
<td>Predicted probabilities for the selected tweet</td>
<td>Pie Chart (bqplot)</td>
<td>Click on each slice to select the model for the specific label</td>
<td>Bar Chart will be updated with selected model weights</td>
</tr>
<tr>
<td>Model weights assigned to the tokens of the selected tweet</td>
<td>Bar Chart (bqplot)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ANALYZE MISCLASSIFICATIONS

Confusion Matrix

Predicted Probabilities

Features & Model Weights

© 2017 Bloomberg Finance L.P. All rights reserved.
ANALYZE MISCLASSIFICATIONS
ANALYZE MISCLASSIFICATIONS
USE LASSO TO FIND DATA ISSUES
USE LASSO TO FIND DATA ISSUES
Purpose of visualization is insight, not pictures and pictures 😊

Use the dashboard to:

- Analyze misclassifications (using confusion matrix)
- Improve model by adding more features (by looking at tweet token weights)
- Fix data issues (using triangle and lasso)
YIELD CURVE PCA
YIELD CURVE

- Bonds have a fixed maturity (1M, 3M, 10Y) and pay coupons

- Examples of bonds – treasury bonds, corporates, municipals etc.

- Yield Curve: Plot of bond yields against maturities

- Adjacent points on the yield curve move together (correlated)
U.S. TREASURY YIELD CURVE

- 11 tenors/maturities
- Typically upward sloping
- Different shapes
  - Pre-crisis
  - Post-crisis
  - Current
YIELD CURVE DYNAMICS

- Yield for each tenor (point on the yield curve) changes every day

- Problem:
  - How to explain/model the changes in the yield curve driven by 11 correlated variables?
  - Any parsimonious representation possible?
PRINCIPAL COMPONENT ANALYSIS (PCA)

- For a dataset of a large number of correlated variables

- PCA can be used to:
  - Reduce dimensionality
  - Retain as much variance in the dataset as possible

- Typically first few (3-5) PCA factors enough to explain almost all the variance
PCA OVER DIFFERENT TIME PERIODS

- PCA factors vary with time periods
- Need tools to quickly select different time periods/intervals and look at PCA factors
- “Interval Selector” in bqplot can be used to:
  - Quickly select different time intervals
  - Perform computations on the selected time interval using callbacks
YIELD CURVE PCA: BEFORE CRISIS

Constant Maturity Treasury Yield Curve (From 06/01/2006 To 06/01/2007)

Yield Curves (12 selected from the above interval)

Variance Explained: 95%

First 3 PCA factors
YIELD CURVE PCA: AFTER CRISIS

Constant Maturity Treasury Yield Curve (From 09/06/2011 To 09/07/2012)

Yield Curves (12 selected from the above interval)

Variance Explained: 97%

First 3 PCA factors
YIELD CURVE PCA: CURRENT

Constant Maturity Treasury Yield Curve (From 08/09/2016 To 08/10/2017)

Yield Curves (12 selected from the above interval)

Variance Explained: 91%

First 3 PCA factors
DATA ANALYSIS OF U.S. ETFS
EXCHANGE TRADED FUNDS (ETF)

- Pooled investment vehicle like mutual funds

- Traded on exchange: buy/sell ETFs just like stocks using brokerage account

- Gain exposure to all corners of the market – stocks, bonds, FX, commodities etc.

- 2000+ U.S. listed ETFs and growing
ETF ATTRIBUTES

- Asset Class (Stocks, Bonds, Currencies etc.)
- Investment Strategy
- Industry/Sector
- Geographical location
- Market Cap
- Dividend Yield
- Expense Ratio
TILE MAP

- Use Tile Map (MarketMap widget in bqplot) to get a global view of all 2000+ ETFs

- Tile Map can be used as a:
  - Heatmap (color each cell by the values of an attribute)
  - Grouping widget (group by an attribute)
  - Selector (select multiple items by clicking on cells)
USE HEATMAP TO GET A GLOBAL PICTURE
SELECT ETFS AND ANALYZE PERFORMANCE

<table>
<thead>
<tr>
<th>Asset Class</th>
<th>Strategy</th>
<th>Group By</th>
<th>Rating Class</th>
<th>Filters</th>
<th>(126 ETFs)</th>
</tr>
</thead>
</table>

- Asset Class: Fixed Income
- Strategy: Fixed or Managed
- Group By: Maturity
- Rating Class: Maturity

Expenses: 0.10%

Selected Interval: 12/31/2016 To 09/10/2017
RESOURCES

Widget libraries used to build the applications:

ipywidgets: https://github.com/jupyter-widgets/ipywidgets
bqplot: https://github.com/bloomberg/bqplot
pythreejs: https://github.com/jovyan/pythreejs
(and other custom widgets)

Tech at Bloomberg: www.TechAtBloomberg.com