Introduction to Data Science with Cortana: Microsoft Azure Machine Learning

Stephen F. Elston
Principle Consultant, Quantia Analytics, LLC
Overview

- Introduction to Azure Machine Learning
- Tour of Azure ML Studio
- Building a first Azure ML experiment
- A forecasting example with Azure ML and R
- Publishing a web service
Why Azure ML?

- Quickly deploy production solutions as web services
- Models run in a highly scalable secure cloud environment
- Powerful, efficient built-in algorithms
- Extensible with, SQL, Python, and R
- Integrated in Cortana stack
Cortana Analytics Suite:
Transform data into intelligent action
The “Distro” for Intelligence
http://microsoft.com/cortanaanalytics

Information Management
- Azure Data Factory
- Azure Data Catalog
- Azure Event Hub

Big Data Stores
- Azure Data Lake
- Azure SQL Data Warehouse

Machine Learning and Analytics
- Azure Machine Learning
- Azure HDInsight (Hadoop)
- Azure Stream Analytics

Dashboards and Visualizations
- Power BI

Personal Digital Assistant
- Cortana

Perceptual Intelligence
- Face, vision
- Speech, text

Business Scenarios
- Recommendations, customer churn, forecasting, etc.

DATA  INTELLIGENCE  ACTION

Business apps
Custom apps
Sensors and devices

People
Automated Systems
Why Open-Source Tools?

- R and Python widely used in data science
- Highly interactive
- Good visualization
- Vast packages (libraries) of utilities and machine learning algorithms
- Excellent development environments
Azure ML Free Tier Account

- Free Tier Account
- Unlimited time, with restricted priority
- Paid account provides full performance
Quick Start Guide to Azure Machine Learning
Azure ML Studio

- Experiments contain workflow
- Experiments constructed of modules
- Experiments in sharable workspace
- Modules transform data, compute models, score models, and evaluate models
- Create custom modules with SQL, R and Python
- Deploy solutions as web services
Azure ML Documentation Resources

Azure ML tutorials and resources:

Azure ML Gallery:

Documentation and examples for each module
Sample Experiments tab in studio
Azure ML Learning Resources

Book, Microsoft Azure Essentials: Azure Machine Learning
http://www.microsoftvirtualacademy.com/ebooks#9780735698178

Data Science in the Cloud Microsoft Azure Machine Learning and R, O’Reilly Media
Azure ML Learning Resources

Data Science and Machine Learning Essentials
With
Stephen Elston and Cynthia Rudin

http://shop.oreilly.com/product/0636920040255.do

https://www.edx.org/course/data-science-machine-learning-essentials-microsoft-dat203x

http://www.amazon.com/Predictive-Analytics-Microsoft-Machine-Learning/dp/1484212010/ref=la_B00NBELJJI_1_1?ie=UTF8&qid=1441060294&sr=1-1
Data Passed from Module to Module in Azure ML Tables

<table>
<thead>
<tr>
<th>Col1, Col2, Col3, \ldots, ColN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Val11, Val12, Val13, \ldots, Val1N</td>
</tr>
<tr>
<td>\ldots, \ldots, \ldots, \ldots, \ldots, \ldots</td>
</tr>
<tr>
<td>ValM1, ValM2, ValM3, \ldots, ValMN</td>
</tr>
</tbody>
</table>

Rectangular table
N Columns - M Rows
Equal length columns
Azure ML Table Data Types

- Numeric: Floating Point
- Numeric: Integer
- Boolean
- String
- Categorical
- Date-time
- Time-Span
- Image
Building a First Model
Building machine learning models

- Define business problem
- Understand data relationships
- Prepare data
- Construct models
- Evaluate models
- Improve models
- (Cross) validate model
- Publish model
Classification

“Science is the systematic classification of experience.”
George Henry Lewes

- Two class and multi-class
- Examples:
  - Species
  - Movie genre
  - Fraud detection
Machine learning workflow

1. Input data
2. Data Transformation
3. Define Model
4. Split Data
5. Train Model
6. Score (prediction)
7. Evaluate Model
Classification

Formally, given training set \( (x_i, y_i) \) for \( i=1...n \), we want to create a classification model \( f \) that can predict label \( y \) for a new \( x \).

\[
f(x) = \text{function(Events Last Year, Oldest Cable)}
\]
Decision Trees For Classification

- If age > 30, then predict y = 1
- If age ≤ 30, then
  - If #classes > 8, then predict y = -1
  - Else, if #classes ≤ 8, then
    - If female, then predict y = 1
    - Else, if male, then predict y = 1

- If Bschool, then
  - If laundry, then predict y = 1
  - Else, if No Bschool, then
    - If No laundry, then predict y = -1
    - Else, if laundry, then predict y = -1

- Otherwise, predict y = 1
Empirical Bayes Model

- P(U) = 99.0%
  - P(U and -) = 4.95%
  - P(U and +) = 94.05%
- P(D) = 1.0%
  - P(D and -) = 0.9%
  - P(D and +) = 0.1%
- P(+|U) = 95.0%
- P(-|U) = 5.0%
- P(+|D) = 10.0%
- P(-|D) = 90.0%
## Metrics for Classification

### Confusion matrix

<table>
<thead>
<tr>
<th></th>
<th>Predicted Positive</th>
<th>Predicted Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Actual Positive</strong></td>
<td><strong>TP</strong></td>
<td><strong>FN</strong></td>
</tr>
<tr>
<td><strong>Actual Negative</strong></td>
<td><strong>FP</strong></td>
<td><strong>TN</strong></td>
</tr>
</tbody>
</table>
Metrics for Classification

- Accuracy = \( \frac{TP + TN}{(TP + TN + FP + FN)} \)
- Precision or positive predictive value = \( \frac{TP}{(TP + FP)} \)
- Recall = \( \frac{TP}{(TP + FN)} \)
- \( F1 = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \)

<table>
<thead>
<tr>
<th></th>
<th>Predicted Positive</th>
<th>Predicted Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Positive</td>
<td>( TP )</td>
<td>( FN )</td>
</tr>
<tr>
<td>Actual Negative</td>
<td>( FP )</td>
<td>( TN )</td>
</tr>
</tbody>
</table>
For a particular False Positive Rate (FPR), what is the True Positive Rate (TPR)?

AUC

Random Guessing
Demo: Classification Example
Regression Model Example with R or Python in Azure ML
Forecasting

• A forecast is a prediction of a future value

• Examples:
  • Inventory levels
  • Utility demand
  • Service requirements
Linear regression

- Need a function that estimates $y$ for a new $x$.
- The simplest is a linear model.

$$f(x_i) = b_0 + b_1 x_i$$

$$f(\text{clicks}_i) = 500K + 200 \cdot \text{clicks}_i$$
Time series trend and seasonal variation

\[ f(x_i) = b_0 + b_1 x_i + b_2 x_i^2 + b_3 x_i^3 \]
Time series trend and seasonal variation
R or Python?

- R and Python are widely used in data science
- Powerful open-source data science tools
- Python tends to be more systematic and faster
- R contains wider range of packages and analytics capabilities
- R support currently deeper in Azure ML
Developing and testing R and Python

- Azure ML is a production environment
- Interactively develop and test in IDE
- Subset data as needed – download as .csv
- IDE has powerful editor and debugger
- Cut and paste code into Execute R/Python Script module to test in Azure ML
- Jupyter notebooks in preview (Python only, R coming)
Debugging R and Python in Azure ML

- Code tested in IDE should run in Azure ML, but........
- If error occurs look at the error.log or output.log
- From R use print() function
- From Python use sys.stderr.write() from sys
myFrame <- maml.mapInputPort(1,2)

source("src/myScript.R")

maml.mapOutputPort("myFrame")

print("Hello world")
Def azureml_main(inFrame1, inFrame2):

    import my_package

    return "myFrame"

print("Hello world")
def azureml_main(frame1):
    # Set graphics backend: Do this first!
    import matplotlib
    matplotlib.use('agg')

    ## Code to create plots

    ## Save figure in a file for output
    fig.savefig('scatter2.png')

    return frame1
Forecasting Regression Example

- Time series regression forecast milk production in the State of California
- In Gallery (R and Python)
  - https://gallery.azureml.net/Experiment/e616740e68c647ba9bbefa663d037df5
  - https://gallery.azureml.net/Experiment/c8c8fe15c4ee470685cc91d5e19c77dc
- On Github (R and Python)
  - https://github.com/Quantia-Analytics/Contana-Data-Science-Example-R
  - https://github.com/Quantia-Analytics/Cortana-Data-Science-Example-Python
Demo: Regression Example
## Metrics for Classification

### Input

<table>
<thead>
<tr>
<th>Col1</th>
<th>Col2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Val11</td>
<td>Val12</td>
</tr>
<tr>
<td>Val21</td>
<td>Val22</td>
</tr>
</tbody>
</table>

### Output

<table>
<thead>
<tr>
<th>Score</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Out1</td>
<td></td>
</tr>
<tr>
<td>Out2</td>
<td></td>
</tr>
</tbody>
</table>

### User Application

### Request

### Response

### Azure ML Service

### End point
Published Input

Input Transformations

Score Trained Model

Output Transformation

Published Output
Demo: Create Web Service
Building machine learning models

- Define business problem
- Understand data relationships
- Prepare data
- Construct models
- Evaluate models
- Improve models
- (Cross) validate model
- Publish model
Why Azure ML?

- Quickly deploy production solutions as web services
- Models run in a highly scalable secure cloud environment
- Powerful, efficient built-in algorithms
- Extensible with, SQL, Python, and R
- Integrated in Cortana stack