Microsoft Azure Machine learning Algorithms

Tomaž KAŠTRUN
@tomaz_tsql
Tomaz.kastrun@gmail.com
http://tomaztsql.wordpress.com

PASS SQL saturday
#494 | VIENNA 2016

Our Sponsors

Microsoft

PASS

PURE STORAGE

SolidQ
Think Big. Move Fast.

DELL Software

PYRAMID ANALYTICS

Profisee®

IDERA

pmOne

UBIDO
AN ACP GROUP COMPANY

redgate
ingeniously simple

ApexSQL

Speaker info

- BI Developer (MSSQL Server, C#, SAS, R, SAP, Py)
- 15+ ys experience MSSQL Server
- 15+ ys experience data analysis and DM
- Working: Spar ICS Österreich, Spar Slovenija
- MCT, MCPT, MCSE SQL Server
- tomaz.kastrun@gmail.com
- @tomaz_tsql
- https://tomaztsql.wordpress.com
- Publishing articles, speaking at SQL events, conferences
- Coffee Lover, Bicycle junkie

Agenda

- Focus on explanation of algorithms available for predictive analytics in Azure Machine Learning service.
- Algorithms
- 1) regression algorithms,
- 2) Two-Class classifications,
- 3) Multi-class classification,
- 4) Clustering

- Explore algorithm
- which algorithm is used and useful for what kind of empirical problem
- which is suitable for particular data-set.

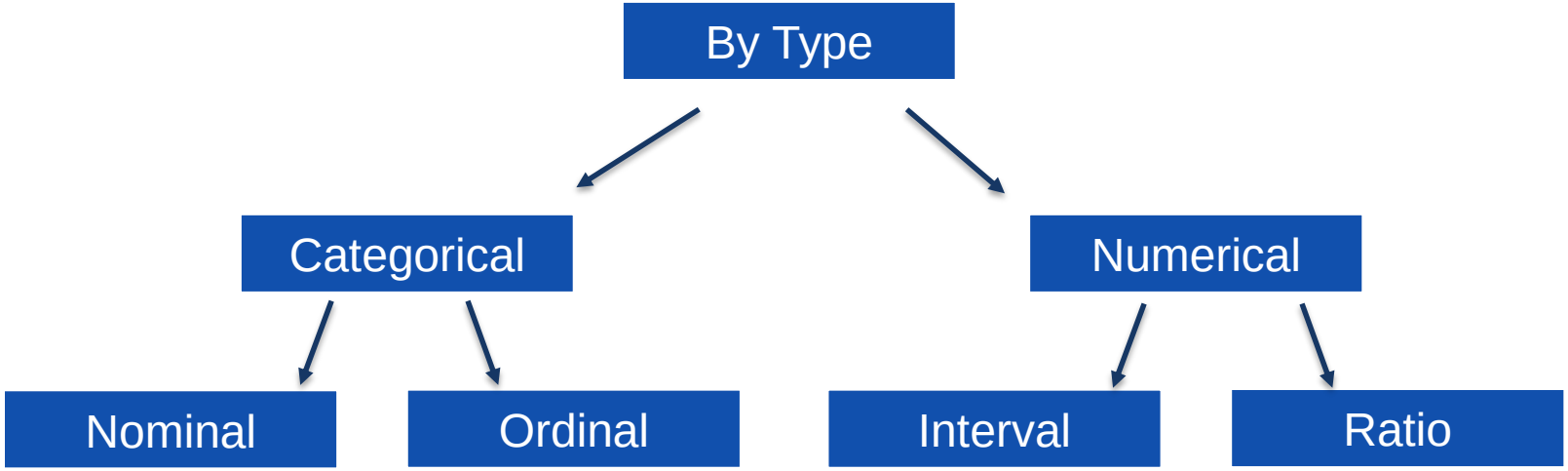
Before start... Why this session?!

- Machine learning is suddenly very popular
- All non-scientists and non-statisticians are now data wranglers and data scientists
- Very easy to accomplish „something“
- No knowledge needed for „something“ that does „something“ and returns „something“

First things first

- Statisticst! (with „something“ you will get „something“)
- Variables (columns/data) by type:
- Variables (columns/data) by input and ouput:
  - Dependent (outcome)
  - Independent (experimental or predictor)

First things „something“



gender (male/ Female), eye colour (blue, green, black,...), marital status (S, M, D,V,...), Buyer | Nonbuyer (0 | 1)

No inherited order; all values are the „same“ and all values equal in representation

education (primary, secondary, high school,...), Salary (<101€,101-200€, 201-300€,..)

Ordering can be applied; values can be compared with =,>,< and recoded with numbers / classes.

Height (173cm, 196cm), Invoice Value (24.4€, 42.6€,..)

Ordering can be applied; values can be compared with =,>,< and expressed 2 is 2x times bigger..

Temperature: Celsius vs. Fahrenheit

For warmup...regression

$$y_i = a + bx_i + e_i$$

$$a + bx$$

y_i = specific y value (dependent variable)

a = intercept

b = slope

x_i = specific x value (independent variable)

e_i = random variance or the residual

continuous predictors	**categorical predictors**
x = a set of continuous data points	x = a set of ***binary/categorical*** data points
a = the value of y when x is 0	a = the value of y when the x is the ***default level***
b = the change in y for one change in x	b = the change in y when x is the ***non-default level***

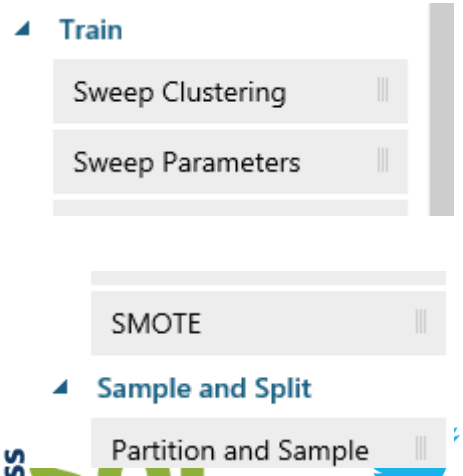
$$log[p/(1-p)]_i = a + bx_i$$

First things second

- Distribution (normal, poission, bernoulli,...)
- Data normalization
- Linear vs. Non-linear problem
- Supervised vs. Unsupervised

When choosing algorithm for ML Azure *keep in mind*:

- Accuracy of algorithm
- Training Time
- Linearity
- Number of parameters in algorithm (module: Parameter sweeping)
- Number of data variables / features
- Data distribution
- Biased Training data (SMOTE)



Sample Data

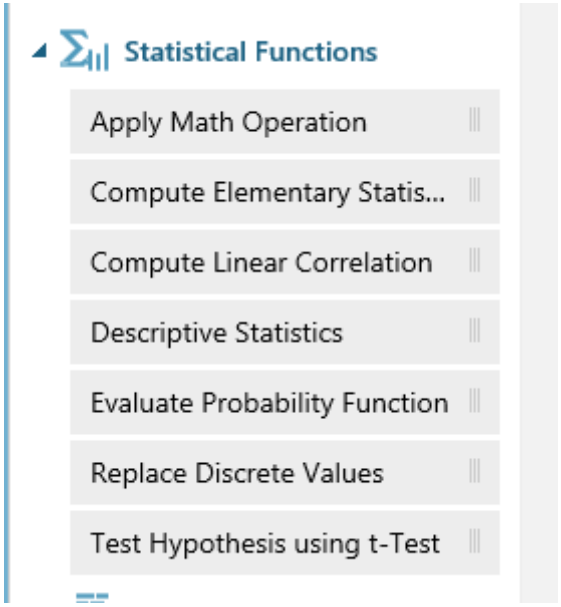
rows: 200 columns: 8

view as

id	Gender	Favorite_ice_cream	Score_video_game	Score_puzzle_game	Buyer	Salary	Salary_recoded
70	0	1	47	57	1	94.470686	5
121	1	2	63	61	0	49.534943	3
86	0	3	58	31	1	89.581054	5
141	0	3	53	56	0	33.229919	2
172	0	2	53	61	0	21.44842	2
113	0	2	63	61	0	2.85406	1
50	0	2	53	61	0	13.096278	1
11	0	2	39	36	1	62.648952	4
84	0	2	58	51	0	32.936928	2
48	0	2	50	51	0	15.021996	1
75	0	2	53	61	0	39.03202	2
60	0	2	63	61	1	68.371854	4
95	0	3	61	71	0	25.211896	2
104	0	3	55	46	1	57.073408	3
38	0	1	31	56	1	64.52152	4
115	0	1	50	56	0	6.542734	1
76	0	3	50	56	0	27.236401	2
195	0	2	58	56	1	79.242704	4
114	0	3	55	61	1	65.905424	4
85	0	2	53	46	0	8.46832	1

1 General (interferance) statistics

Statistical Functions



Apply Math Operation -> Applies a mathematical operation to column values
Compute Elementary Statistics -> Calculates specified summary statistics for selected dataset columns
Compute Linear Correlation -> Calculates the linear correlation between column values in a dataset
Descriptive Statistics -> Generates a basic descriptive statistics report for the columns in a dataset
Evaluate Probability Function -> Fits a specified probability distribution function to a dataset
Replace Discrete Values -> Replaces discrete values from one column with numeric values based on another column
Test Hypothesis Using t-Test -> Compares means from two datasets using a t-test

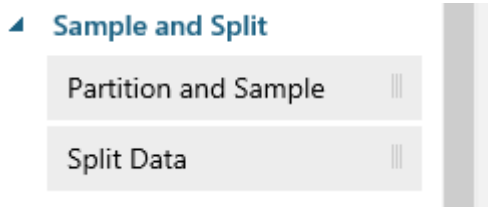
Source: https://msdn.microsoft.com/en-us/library/azure/dn905867.aspx

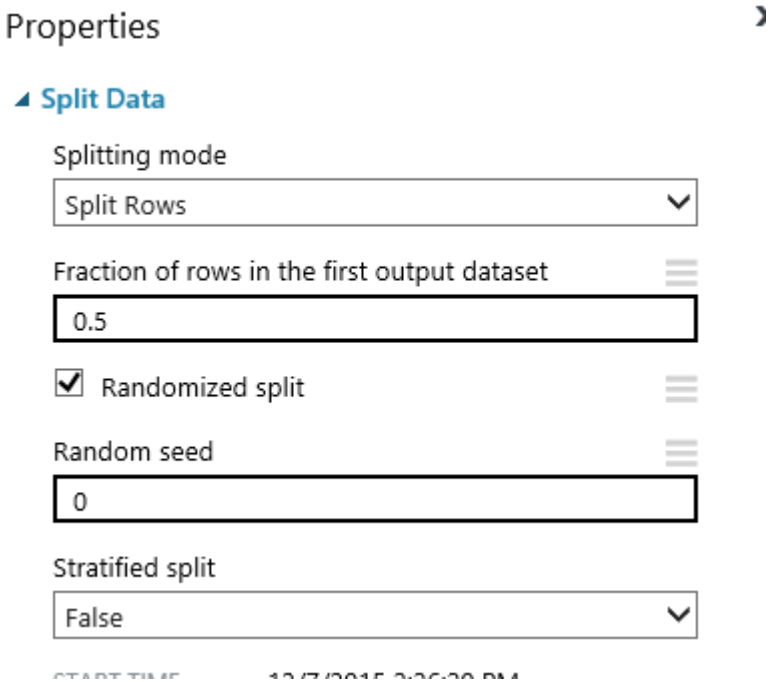
1 General Statistics

Short DEMO

1 Spliting data

- Spliting data





Splitting Mode:
- Split rows
- Recommender split
- Regular / Relative expression

Random seed
Stratified split

Jackknife, cross validation, n-fold cross validation, ...

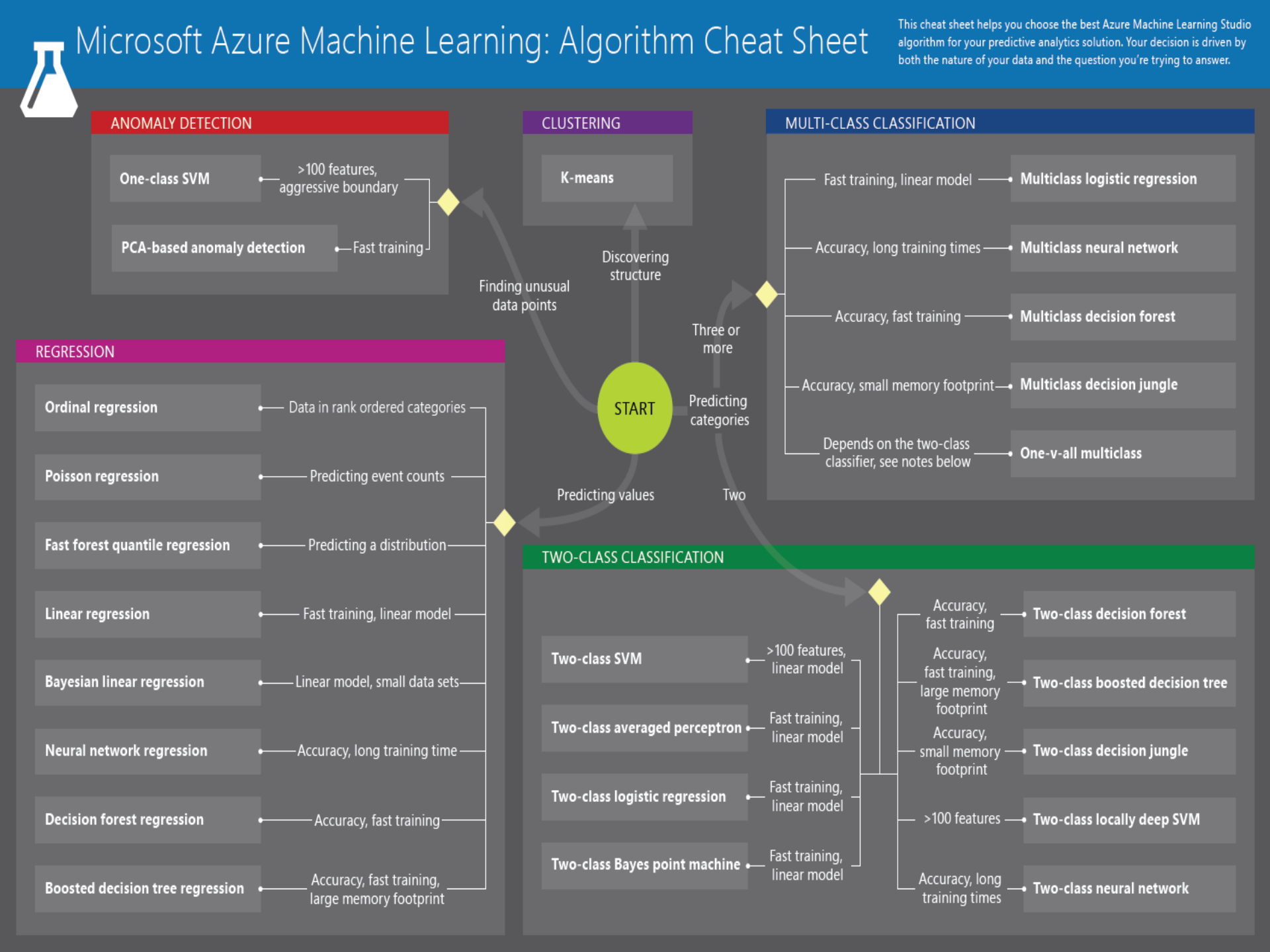
1 Spliting data

Short DEMO

1 Initializing Model

Microsoft Azure Machine Learning: Algorithm Cheat Sheet

This cheat sheet helps you choose the best Azure Machine Learning Studio algorithm for your predictive analytics solution. Your decision is driven by both the nature of your data and the question you're trying to answer.



1 Initializing Model

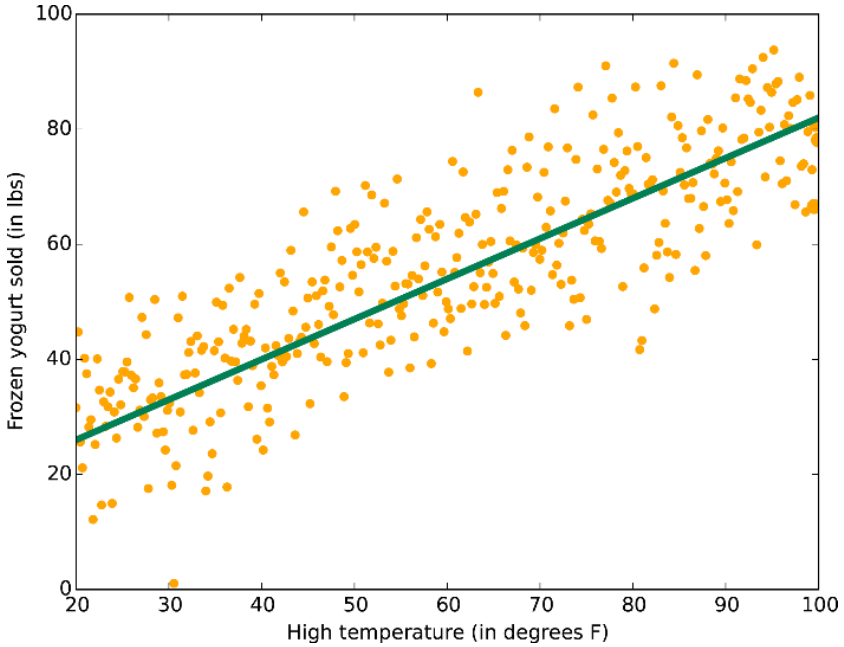
Making list of algorithms more transparent

	Regresssion	Classification	
		Two-class	Multiclass
Average Perceptron		✓	
Bayes Point Machine		✓	
Decision Forest	✓	✓	✓
Decision Jungle		✓	✓
Decision Tree	✓	✓	
Fast Forest	✓		
Linear Regression	✓		
Bayes Linear Regression	✓		
Log Regression		✓	✓
Neural Network	✓	✓	✓
Ordinal Regression	✓		
Poisson Regression	✓		
SVM		✓	
SVM Deep Support		✓	

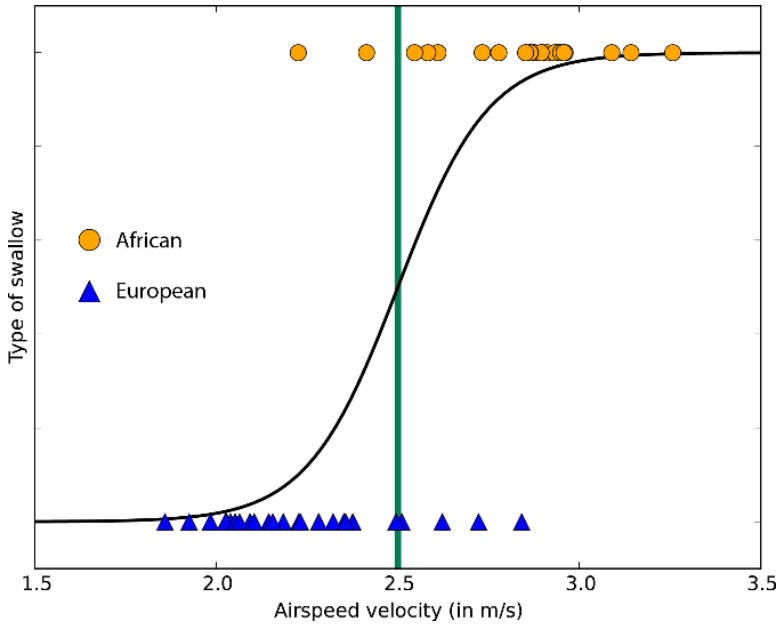
Algorithms in Theory

Regression

Linear and Logistic



Azure ML: Linear Regression

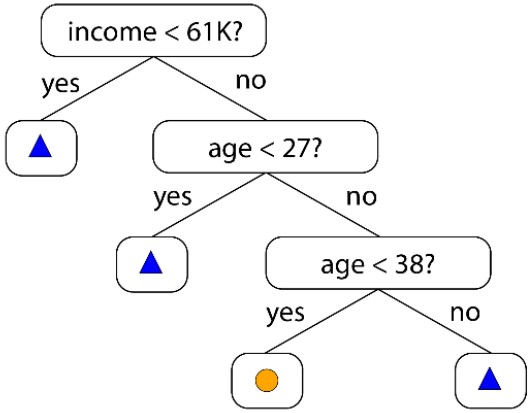


Azure ML: Two-class Classification Logistic Regression
Multiclass classification Logistic Regression

Algorithms in Theory

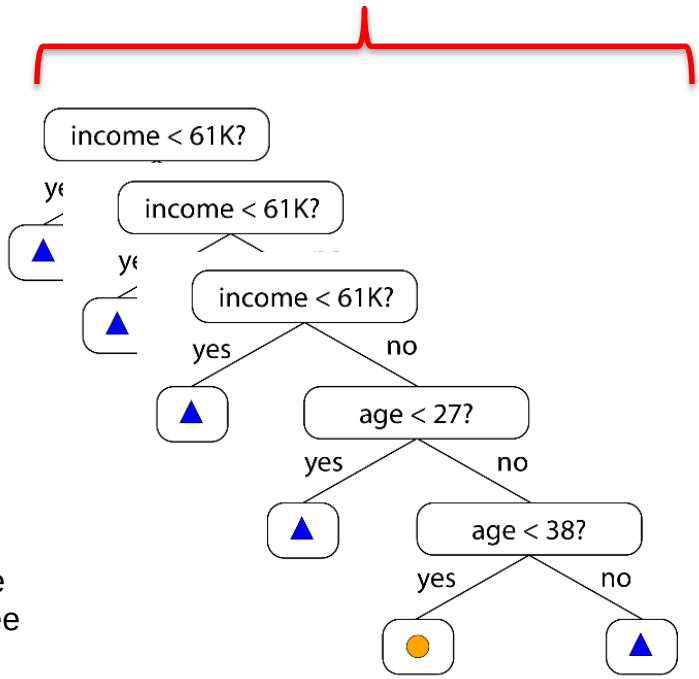
Decision Tree, Decision Forests, Decision Jungles

Decision tree



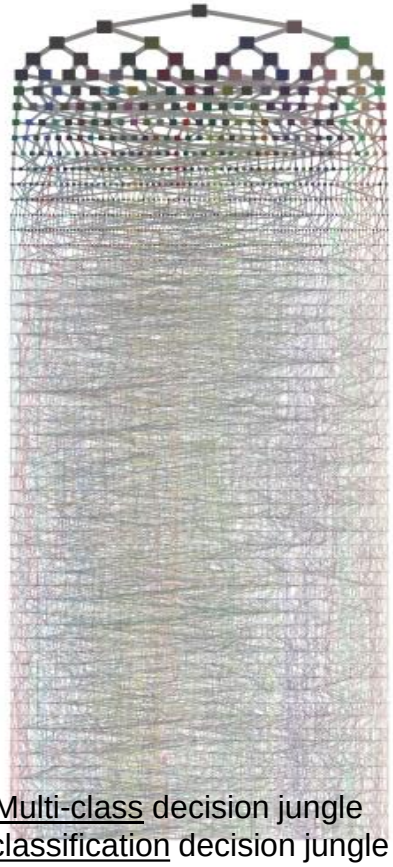
Azure ML: Regression boosted decision tree
Two-class classification boosted decision tree

Decision Forest



Azure ML: Regression decision forrest
Two-class classification decision forrest
Multi classification decision forrest

Decision Jungle

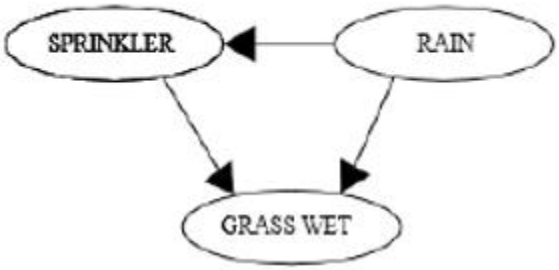


Azure ML: Multi-class decision jungle
Two-class classification decision jungle

Algorithms in Theory

Naive Bayes

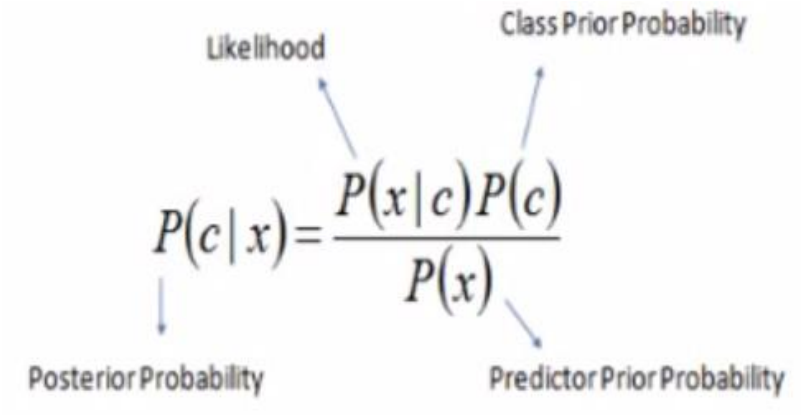
RAIN	SPRINKLER T	SPRINKLER F
F	0.4	0.6
T	0.01	0.99



RAIN T	RAIN F
0.2	0.8

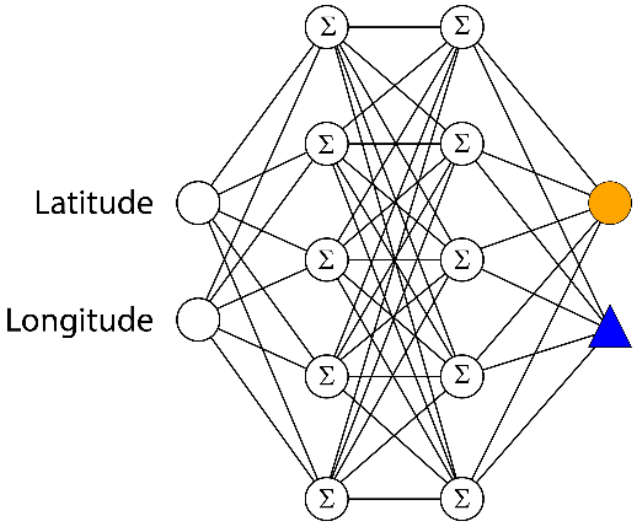
SPRINKLER	RAIN	GRASS WET T	GRASS WET F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

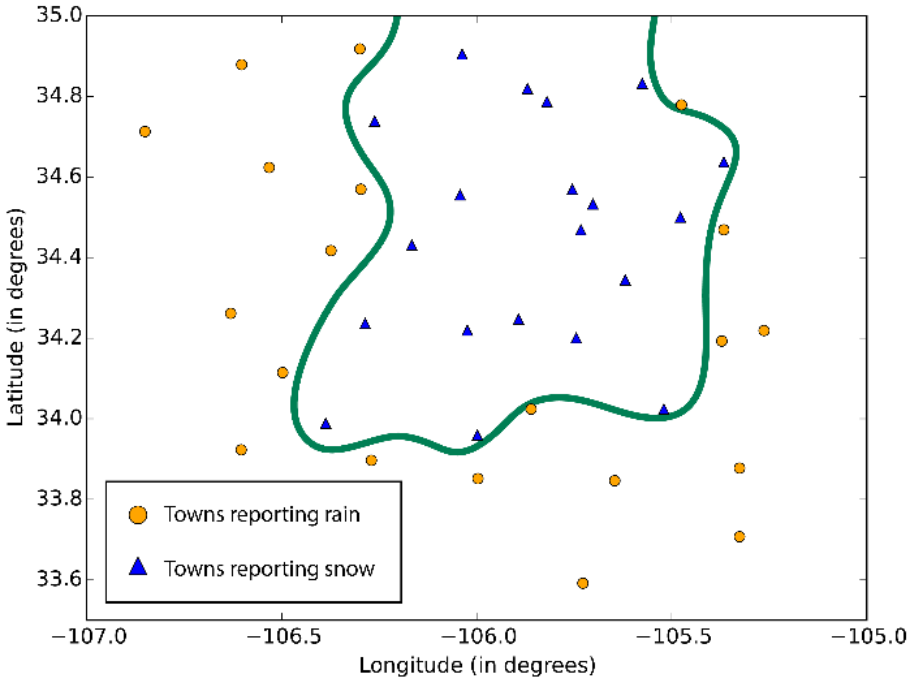
Azure ML: Regression Bayes linear
Two Class classification Bayes' point machine



Algorithms in Theory

Neural networks and perceptrons



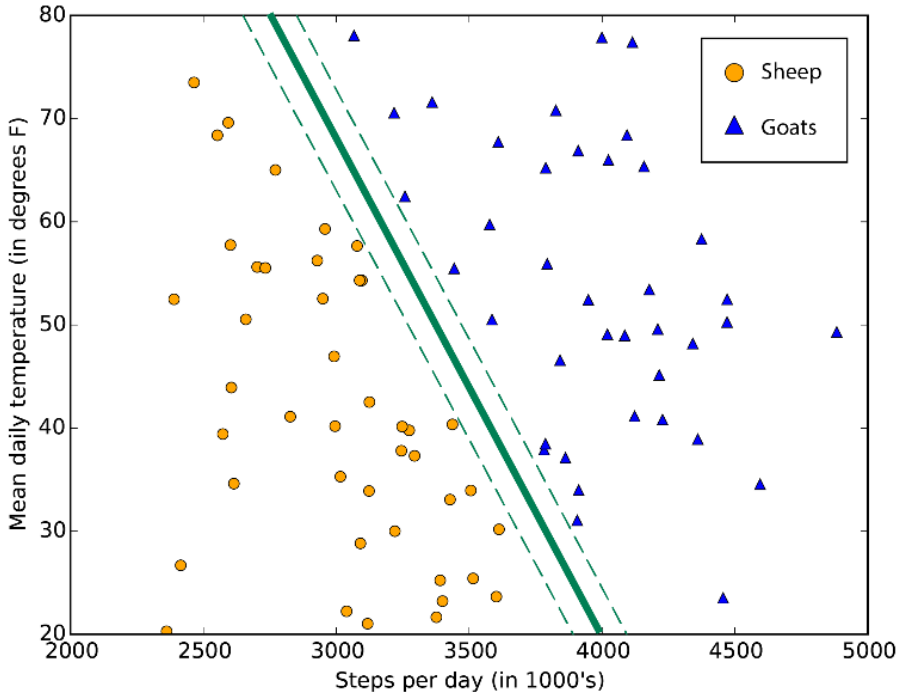


Azure ML: Regression Neural networks
Two Class classification Neural networks
Multi Class classification Neural networks

Two Class Classification averaged perceptrons

Algorithms in Theory

SVM



Azure ML: Two Class classification SVM
Two Class classification locally deep SVM

Anomaly detection SVM

1 Regression Algorithms

- Regression is method for estimating relations among parameters/varibles.
- Linear vs. Logistic (linear combination of parameters vs. Logistic combination of parameters)
- Typical Problem would be predicting Y ; a numeric value.
- Typical Azure Algorithms
  - Boosted Decision Tree Regression
  - Decision Forest Regression
  - Linear Regression
  - Bayesian Linear Regression

1 Regression Algorithms Parameters

Linear Regression

Solution method: Ordinary Least Squares

L2 regularization weight: 0.001

Include intercept term

Random number seed:

Allow unknown categ...

Bayesian Linear Regression

Regularization weight: 1

Allow unknown categ...

Decision Forest Regression

Resampling method: Bagging

Create trainer mode: Single Parameter

Number of decision trees: 8

Maximum depth of the d...: 32

Number of random splits...: 128

Minimum number of sam...: 1

Allow unknown value...

Boosted Decision Tree Regressi...

Create trainer mode: Single Parameter

Maximum number of leav...: 20

Minimum number of sam...: 10

Learning rate: 0.2

Total number of trees con...: 100

Random number seed:

Allow unknown categ...

Neural Network Regression

Create trainer mode: Single Parameter

Hidden layer specification: Fully-connected case

Number of hidden nodes: 100

Learning rate: 0.005

Number of learning iterat...: 100

The initial learning weight...: 0.1

The momentum: 0

The type of normalizer: Min-Max normalizer

Shuffle examples

Random number seed:

Allow unknown categ...

1 Regression Algorithms

Short DEMO

Evaluating Regression Algorithms

Mean Absolute Error:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i|$$

Metrics to measure how close predictions are to eventual outcomes

Root Mean Square Error:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(f_i - y_i)^2}{n}}$$

Metrics of differences between predicted values and actual values.

Relative square Error:

$$RSE = \frac{\sum_{i=1}^{n}(f_i - y_i)^2}{\sum_{i=1}^{n}(\overline{y} - y_i)^2}$$

Coeff. of Determination:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

Summarization of regression model how well fits a statistical model; R^2 = 1 model is perfect, respectively

Evaluating Regression Algorithms

Predicting Salary

03_Regression_algorithm_R

03_Regression_algorithm_R › Add Rows › Results dataset

rows 5 columns 6

view as

Algorithm	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Relative Squared Error	Coefficient of Determination
Linear Regression	12.29696	14.58386	0.489833	0.26734	0.73266
Decision Forest Regression	14.426837	17.068578	0.574674	0.366196	0.633804
Bayesian Linear Regression	12.627958	14.970776	0.503018	0.281714	0.718286
Boosted Decision Tree Regression	14.829949	17.810291	0.590732	0.398714	0.601286
Neural Networl Regression	14.588252	17.217506	0.581104	0.372615	0.627385

Comparison of Regression Algorithms

Regression Algorithm	Accuracy	Training time	Linearity	Customization	Predicting Variable	Type of independant variable(s)	Data Quantity
linear	Good	Fast	Excellent	Good	Interval	Any	small to big
Bayesian linear	Good	Fast	Excellent	Moderate	Interval	Any	big
decision forest	Excellent	Moderate	Good	Good	Interval	Any	
boosted decision tree	Excellent	Fast	Good	Good	Interval	Any	big
fast forest quantile	Excellent	Moderate	Moderate	Excellent	Distribution (Interval)	Any	
neural network	Excellent	Slow	Moderate	Excellent	Interval	Any	smaller
Poisson	Good	Moderate	Excellent* (log linear)	Good	Interval (counts)	Any	small to big
ordinal	Good	Moderate	Excellent	None	Ordinal (order)	Any	small to big

Scale:

Excellent	Good	Moderate
Fast	Moderate	Slow

2 Two-class Classification

- Creates classification estimates for label / prediction variable with dichotomious values
- Typical Problem would be predicting a binary class for label variable
- Typical Azure Algorithms
  - Boosted Decision Tree two-class
  - Decision Forest two-class
  - Decision Jungle two-class
  - Logistic Regression two-class
  - Neural Network two-class
  - Averaged Perceptron two-class
  - SVM two-class

2 Two-class Classification parameters

Two-Class Boosted Decision Tree

Create trainer mode: Single Parameter

Maximum number of leav...: 20

Minimum number of sam...: 10

Learning rate: 0.2

Number of trees construc...: 100

Random number seed:

Two-Class Decision Jungle

Resampling method: Bagging

Create trainer mode: Single Parameter

Number of decision DAGs: 8

Maximum depth of the d...: 32

Maximum width of the de...: 128

Number of optimization s...: 2048

Two-Class Averaged Perceptron

Create trainer mode: Single Parameter

Learning rate: 1

Maximum number of iter...: 10

Random number seed:

Allow unknown categ...

Two-Class Bayes Point Machine

Number of training iterati...: 30

Include bias

Allow unknown value...

Two-Class Support Vector Mac...

Create trainer mode: Single Parameter

Number of iterations: 1

Lambda: 0.001

Normalize features

Project to the unit-sp...

Random number seed:

Allow unknown categ...

Two-Class Logistic Regression

Create trainer mode: Single Parameter

Optimization tolerance: 1E-07

L1 regularization weight: 1

L2 regularization weight: 1

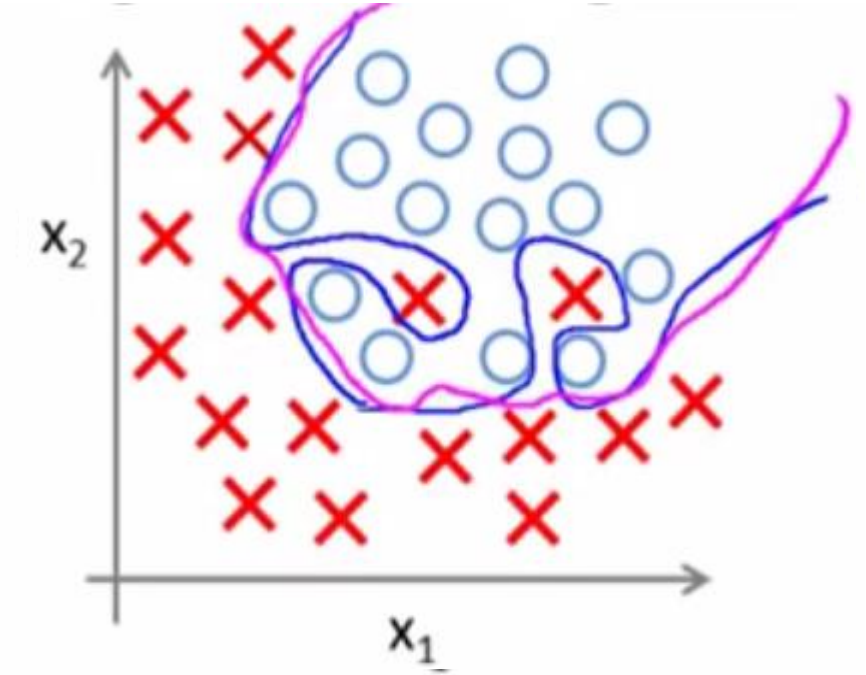
Memory size for L-BFGS: 20

Random number seed:

Allow unknown categ...

Regularization weight

- Used for avoiding overfitting.
- L1, L2 penalized estimation methods shrink the estimates of regre. coefficient towards zero in relation to maximize likelihood of estimates.
- L1 - for sparse, high-dimensional model
- L2 – for dense (or smaller) model and computationally efficient

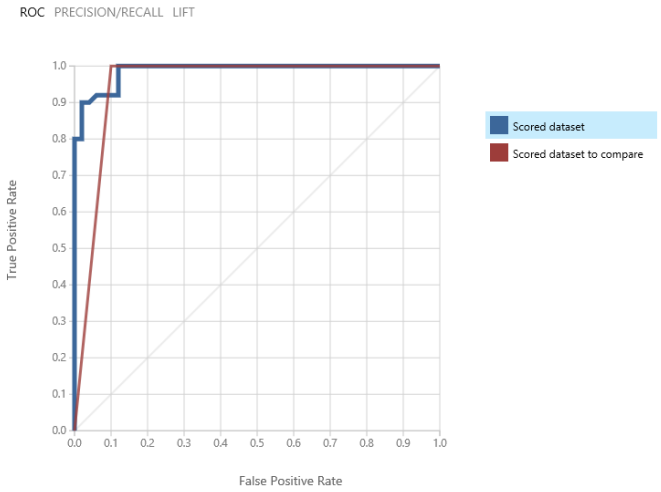


2 Two-class Classification

Short DEMO

2 Evaluating two-class Classificatio

ROC (AUC) Curve / Precision / Lift Chart



AUC/ROC:
<= 0.5 -- ☹☹
0.5 – 0.6 -- ☹
0.6 – 0.7 -- 😐
0.7 – 0.8 -- ☺
0.8 – 0.9 -- ☺☺
0.9 – 1 -- WTF?

Classification Matrix / Confussion matrix / Metrics

True Positive	False Negative	Accuracy	Precision	Threshold	AUC
46	4	0.900	0.885	0.5	0.988

False Positive	True Negative	Recall	F1 Score
6	44	0.920	0.902

Positive Label	Negative Label
1	0

2 Evaluating two-class Classificatio

Metrics

True Positive	False Negative	Accuracy	Precision	Threshold	AUC
46	4	0.900	0.885	0.5	0.988

False Positive	True Negative	Recall	F1 Score
6	44	0.920	0.902

Positive Label	Negative Label
1	0

True Positive (TP) – correctly identified: Buyes is classified as Buyer
False Positive (FP) – Incorrectly identified: Buyes is classified as non-buyer
True Negative (TN) - correctly identified: Non-buyes is classified as non-buyer
False Negative (FN) - Incorrectly identified: Non-buyes is classified as buyer

Accuracy (TP + TN) / (TP + TN + FP + FN) – Proportion of correctly classified
Precision TP / (TP+FP) – Proportion of positive cases classified correctly
Sensitivity* TP / (TP + TN) - Proportion of actual positive cases classified correctly
Score 2TP / (2TP + FP + FN) – Harmonic mean of precision and Sensitivity

Sensitivity is also known as Recall

Comparison of Two-class Classification Algorithms

Two-class classification	Accuracy	Training time	Linearity	Customization	Predicting Variable	Type of independant variable(s)	Data Quantity
logistic regression	Good	Fast	Excellent	Good	dichotomous / binary	Any	small-big
decision forest	Excellent	Moderate	Good	Good	dichotomous / binary	Any	small-big
decision jungle	Excellent	Moderate	Good	Good	dichotomous / binary	Any	big
boosted decision tree	Excellent	Moderate	Good	Good	dichotomous / binary	Any	big
neural network	Excellent	Slow	Moderate	Excellent	dichotomous / binary	Any	
averaged perceptron	Good	Moderate	Excellent	Moderate	dichotomous / binary	Any	
support vector machine	Excellent	Moderate	Excellent	Good	dichotomous / binary	Any	big
locally deep support vector machine	Good	Slow	Good	Excellent	dichotomous / binary	Any	big
Bayes' point machine	Moderate	Moderate	Excellent	Moderate	dichotomous / binary	Any	

Scale:

Excellent	Good	Moderate
Fast	Moderate	Slow

3 Multi-class Classification

- Creates classification estimates for label / prediction variable with 2+ classes
- Decision trees vs. Logistic Regression vs. Neural Network
- Typical Problem would be predicting a class for label variable
- Typical Azure Algorithms
  - Decision Forest Multiclass
  - Decision Jungle Multiclass
  - Logistic Regression Multiclass
  - Neural Network Multiclass

3 Multi-class Classification parameters

Multiclass Decision Forest

Resampling method: Bagging

Create trainer mode: Single Parameter

Number of decision trees: 8

Maximum depth of the d...: 32

Number of random splits...: 128

Minimum number of sam...: 1

Allow unknown value...

Multiclass Decision Jungle

Resampling method: Bagging

Create trainer mode: Single Parameter

Number of decision DAGs: 8

Maximum depth of the d...: 32

Maximum width of the de...: 128

Number of optimization s...: 2048

Allow unknown value...

Multiclass Logistic Regression

Create trainer mode: Single Parameter

Optimization tolerance: 1E-07

L1 regularization weight: 1

L2 regularization weight: 1

Memory size for L-BFGS: 20

Random number seed:

Allow unknown categ...

3 Multi-class Classification

Short DEMO

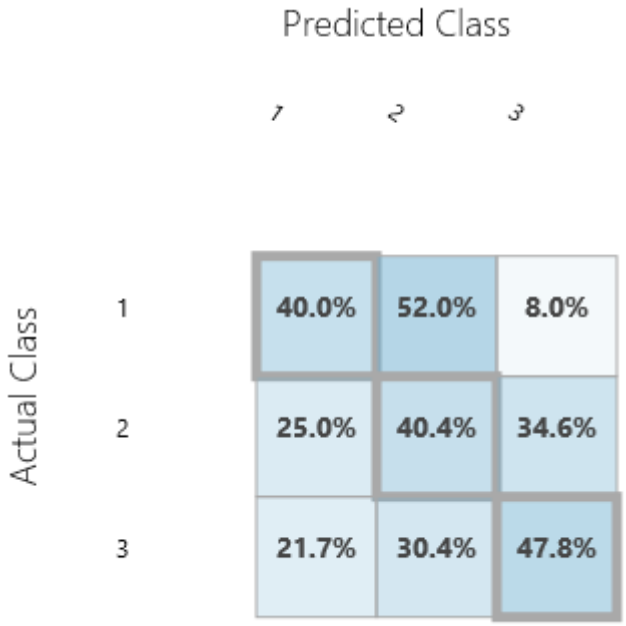
3 Evaluating multi-class Classificati

Metrics

Metrics

Overall accuracy	0.42
Average accuracy	0.613333
Micro-averaged precision	0.42
Macro-averaged precision	0.408059
Micro-averaged recall	0.42
Macro-averaged recall	0.427369

Confusion Matrix



3 Evaluating multi-class Classificati

rows 12 columns 7

view as

Class	Predicted as "1"	Predicted as "2"	Predicted as "3"	Average Log Loss	Precision	Recall
1	10	13	2	4.614802	0.357143	0.4
2	13	21	18	5.49525	0.512195	0.403846
3	5	7	11	2.129248	0.354839	0.478261
1	4	14	7	2.766719	0.444444	0.16
2	2	25	25	2.397737	0.581395	0.480769
3	3	4	16	0.980574	0.333333	0.695652
1	2	19	4	1.354347	0.5	0.08
2	2	35	15	0.842228	0.530303	0.673077
3	0	12	11	0.92375	0.366667	0.478261
1	3	19	3	1.32618	0.375	0.12
2	4	34	14	0.760755	0.548387	0.653846
3	1	9	13	0.951068	0.433333	0.565217

3 Comparison of Multi-class Classificati

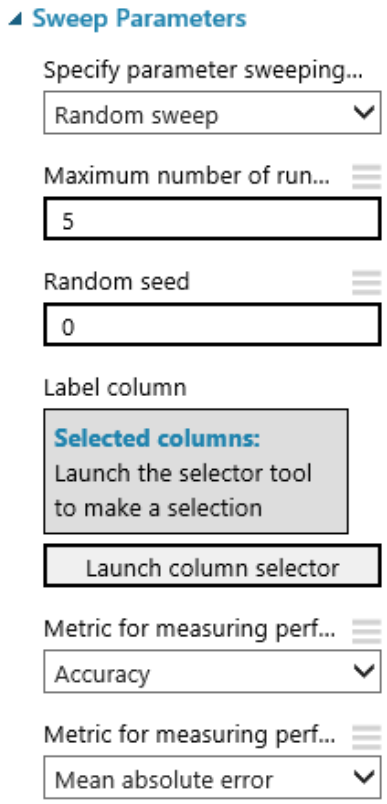
Multi-class classification	Accuracy	Training time	Linearity	Customization	Predicting Variable	Type of independant variable(s)	Data Quantity
logistic regression	Good	Fast	Excellent	Good	Nominal / ordinal (with 2+ classes)	any	small-big
decision forest	Excellent	Moderate	Good	Good	Nominal / ordinal (with 2+ classes)	any	big
decision jungle	Excellent	Moderate	Good	Good	Nominal / ordinal (with 2+ classes)	any	big
neural network	Excellent	Slow	Moderate	Excellent	Nominal / ordinal (with 2+ classes)	any	small

Scale:

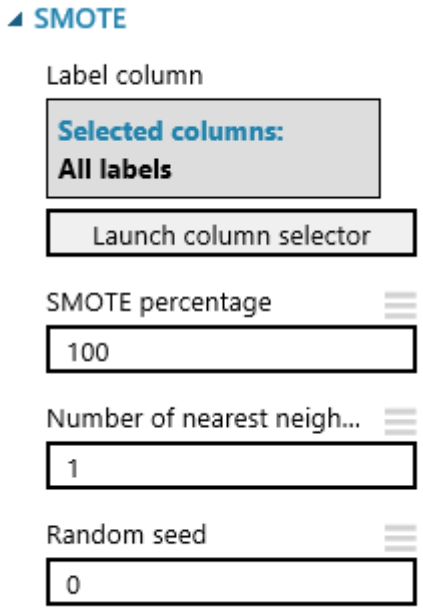
Excellent	Good	Moderate
Fast	Moderate	Slow

4 Using Sweeping and SMOTE

Sweep helps automatically finds the best parameter setting.



SMOTE helps with imbalanded datasets.



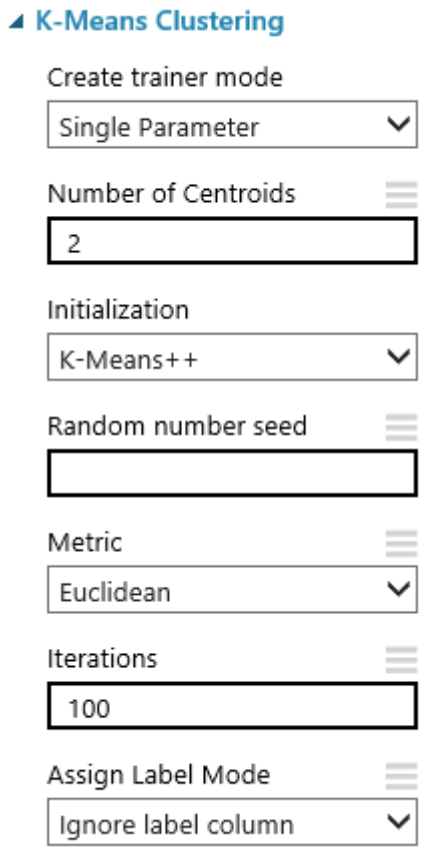
4 Using Sweeping and SMOTE

Short DEMO

5 Clustering

- Assigns value to a given centroid/cluster based on similarity/dissimilarity of values
- Unsupervised machine learning algorithm
- Typical Azure Algorithms
  - K-means Clustering

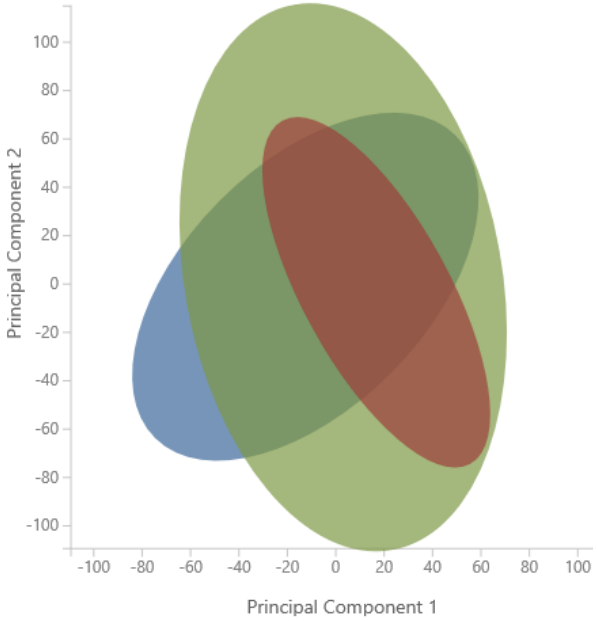
5 Clustering



5 Clustering

Short DEMO

5 Evaluating Clustering



Cluster 0
Cluster 2
Cluster 1

view as

Gender	Favorite_ice_cream	Score_video_game	Score_puzzle_game	Buyer	Salary_recoded	Assignments	DistancesToClusterCenter no.0	DistancesToCluster no.1
0	3	61	66	1	5	0	4.516345	34.428277
0	2	58	51	0	2	2	12.023479	22.491586
0	2	39	46	1	4	1	25.931697	5.941905
0	1	42	26	1	4	1	40.539523	14.764887
0	2	58	51	0	2	2	12.023479	22.491586
1	2	51	58	1	4	2	9.147412	21.880151
0	2	50	51	0	1	2	14.858878	16.057023
1	1	53	37	0	2	2	26.597987	15.125624
1	2	61	56	0	2	0	7.450402	27.673597

Key takeaways

- Knowing your data-sets and data-types
- Knowing group of algorithms
- Correct algorithms for suitable data
- Use Azure ML cheat-sheet diagram when in doubt (https://azure.microsoft.com/en-us/documentation/articles/machine-learning-algorithm-cheat-sheet/)
- A-Z list of Modules (https://msdn.microsoft.com/en-us/library/azure/dn906033.aspx)
- Play with algorithms
- Always double check your Azure ML Algorithms in R or Python (!!)

How did you like it?

Please give feedback

to the event:
http://www.sqlsaturday.com/494/eventeval.aspx

to me as a speaker:
http://www.sqlsaturday.com/494/sessions/sessionevaluation.aspx

THANKS!