Using GLRM on H2O.ai

to Predict Customer Value

Corinne Powers, PhD

About Me

Sr. Data Scientist at Retina AI

Math, Optimization and Signals Modeling Background

R and Python

New to LA Community (8 months)

This Talk

Business Value

Modeling & Data

R Usage in H2O.ai

Motivation

- You sell Animal Cookies through your website

- Customers may buy animal packs or subscribe monthly

- You track session data of site visits, even if they don't purchase

How do we **infer** behavior to **target** customers who could be valuable?

Clustering!

Big Picture: Seek people who "look" similar to high-value customers.

Supervised Learning

Decision tree

K-nearest neighbors

Unsupervised Learning

K-Means clustering

But where can this fail? Any why?

Mixed Data Types

	Numerical	Categorical				Boolean
	Total Spend	*Device Type*	*Acquisition Channel*	*Favorite Shape*	*Favorite Flavor*	*Subscriber*
Customer A	$$	Mobile	Facebook	Tiger	Sprinkles	No
Customer B	$	Mobile	Google	Elephant	Chocolate	No
Customer C	$$$$$	Desktop	Bing	Lion	Sprinkles	Yes
Customer D	$$$	Tablet	Google	Elephant	Sprinkles	No

Semi Supervised Learning

	Numerical		**Categorical**			**Boolean**
	Total Spend	*Device Type*	*Acquisition Channel*	*Favorite Shape*	*Favorite Flavor*	*Subscriber*
Customer A	$$	Mobile	Facebook	Tiger	Sprinkles	No
Customer B	$	Mobile	Google	Elephant	Chocolate	No
Customer C	$$$$$	Desktop	Bing	Lion	Sprinkles	Yes
Customer D		Tablet	Google	Elephant	Sprinkles	

Missing Data

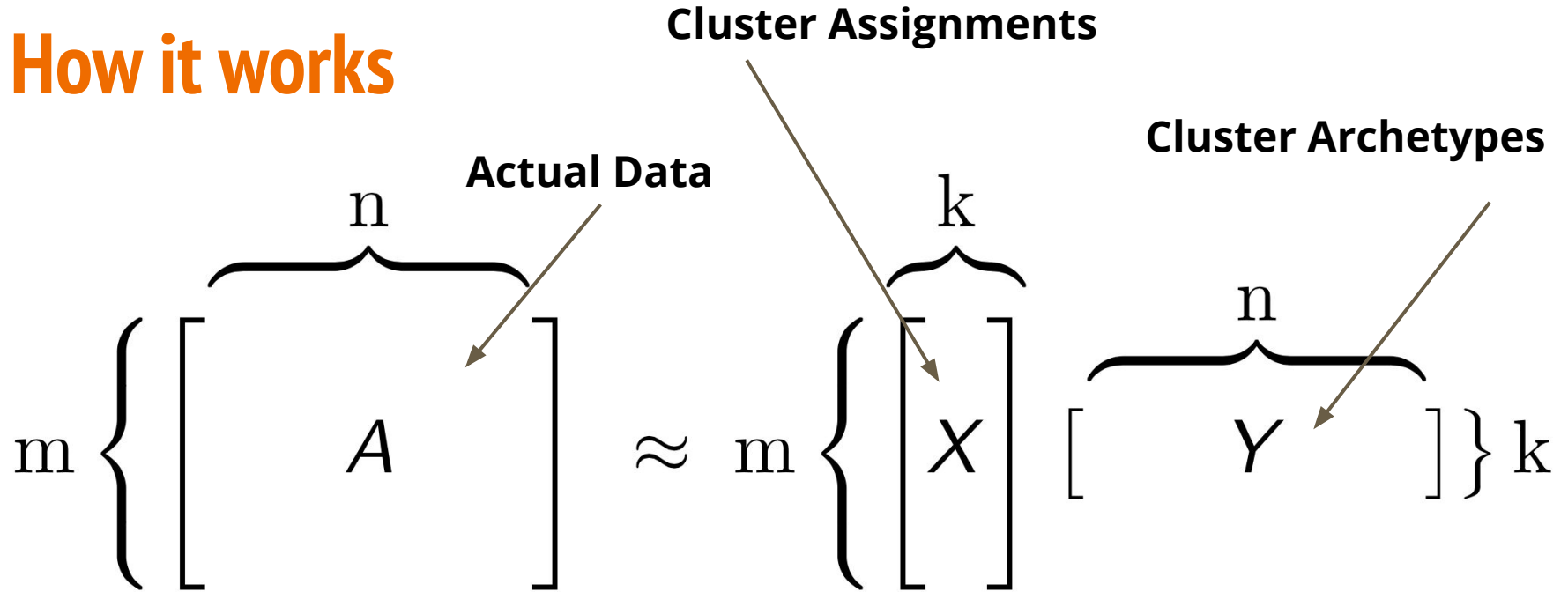|	**Numerical**		**Categorical**			**Boolean**
	Total Spend	*Device Type*	*Acquisition Channel*	*Favorite Shape*	*Favorite Flavor*	*Subscriber*
Customer A	$$	Mobile	Facebook	Tiger	Sprinkles	No
Customer B	$	Mobile	Google			No
Customer C	$$$$$	Desktop	Bing	Lion		Yes
Customer D		Tablet	Google	Elephant	Sprinkles	

Generalized Low Rank Models (GLRMs)

Capabilities: handle missing data, handle weird types of data, you pick feature importance

Special Cases: SVM, K-means clustering, PCA

Business Value: Build Personas and Predict $$$, Consumer Preferences, etc.

https://web.stanford.edu/~boyd/papers/pdf/glrm.pdf

How it works

Actual Data

Cluster Assignments

Cluster Archetypes

$$m \left\{ \left[\quad A \quad \right] \approx m \left\{ \left[X \right] \left[\quad Y \quad \right] \right\} k \right.$$

with braces labeling n (columns of A), k (columns of X), and n (columns of Y).

http://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science/glrm.html

GLRM parameters on H2O.ai

```r
1  library(h2o)
2  h2o.init()
```

```r
1  # GLRM parameters
2  glrm_loss <- "Quadratic"       # Least squares
3  glrm_regx <- "UnitOneSparse"   # Assign to one cluster
4  glrm_gammax <- 1.0
5  glrm_gammay <- 1.0
6  glrm_regy <- "NonNegative"     # Interpretable cluster
7  glrm_k <- 8                    # Number of clusters
```
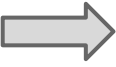
https://cran.r-project.org/web/packages/h2o/index.html

Train a GLRM

```r
visits.glrm <- h2o.glrm(
  training_frame = visits.train,
  cols = union(glrm_features, glrm_responses),
  k = glrm_k,
  loss = glrm_loss,
  regularization_x = glrm_regx, regularization_y = glrm_regy,
  gamma_x = glrm_gammax, gamma_y = glrm_gammay,
  transform = "NONE",
  max_iterations = 1000, max_updates = 2000,
  svd_method = "GramSVD", min_step_size = 1e-6,
  ignore_const_cols = FALSE, seed = 10)
```

Impute Labels via Clusters

```
1   # impute missing values
2   visits.predict <- h2o.reconstruct(visits.glrm,
3                       visits.train[,union(glrm_features, glrm_responses)])
```

	Arch1	Arch2	Arch3	Arch4	Arch5	Arch6	Arch7	Arch8
[1,]	0	0	0	0	0	0	1	0
[2,]	1	0	0	0	0	0	0	0
[3,]	0	0	0	0	1	0	0	0
[4,]	0	0	0	0	1	0	0	0
[5,]	0	0	0	0	0	0	1	0

	reconstr_revenue
1	104.7450
2	110.5066
3	108.6246
4	108.6246
5	104.7450

Target Valuable Customers

- GLRM to create Personas that are **descriptive** and **predictive**

- Match customers to closest Persona to make behavioral inferences

- Even with little data, can match using subset of features as needed

Acknowledgements

Adam Brownell, Emad Hasan, Brad Ito (Retina AI)

- Craft Customer Value business cases

- Support an H2O.ai environment in daily workflow

- Feedback from earlier talks

Anqi Fu (Stanford University)

- Work on GLRM tools for H2O.ai in R

Questions?

corinne@retina.ai

GLRM paper: https://web.stanford.edu/~boyd/papers/pdf/glrm.pdf

H2O.ai Cran: https://cran.r-project.org/web/packages/h2o/index.html

GLRM on H2O.ai: http://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science/glrm.html