Session # C8

EHR Cluster Analysis: Maximizing Patient Care

- Jessica Goodman, PhD, Postdoc Fellow
- Angela Lamson, PhD, Professor and Associate Dean for Research

CFHA Annual Conference
October 17-19, 2019 • Denver, Colorado

CFHA COLLABORATIVE FAMILY HEALTHCARE ASSOCIATION

Faculty Disclosure

The presenters of this session have NOT had any relevant financial relationships during the past 12 months.

CFHA COLLABORATIVE FAMILY HEALTHCARE ASSOCIATION

Conference Resources

Slides and handouts shared by our conference presenters are available on the CFHA website at https://www.cfha.net/page/Resources_2019 and on the conference mobile app.

CFHA COLLABORATIVE FAMILY HEALTHCARE ASSOCIATION

Learning Objectives

At the conclusion of this session, the participant will be able to:

- Describe what a machine learning clustering algorithm can do with a large dataset such as the EHR
- Identify applications for clustering at your own site
- Determine next-steps in cluster analysis process at own site

CFHA COLLABORATIVE FAMILY HEALTHCARE ASSOCIATION

Bibliography / Reference

1. Denaxas, S. C., Asselbergs, F. W., & Moore, J. H. (2016). The tip of the iceberg: challenges of accessing hospital electronic health record data for biological data mining. •
2. Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. (2016). Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Scientific reports, 6, 26094.
3. Nelson, R., & Staggers, N. (2016). Health informatics: An interprofessional approach. Elsevier Health Sciences.
4. Tomar, D., & Agarwal, S. (2013). A survey on Data Mining approaches for Healthcare. International Journal of Bio-Science and Bio-Technology, 5(5), 241-266.
5. Vayena, E. & Blasimme, A. (2017). Bioethical Inquiry, 14, 501-513. https://doi.org/10.1007/s11673-017-9809-6

CFHA COLLABORATIVE FAMILY HEALTHCARE ASSOCIATION

Learning Assessment

- A learning assessment is required for CE credit.
- A question and answer period will be conducted at the end of this presentation.

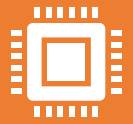
CFHA COLLABORATIVE FAMILY HEALTHCARE ASSOCIATION

Introductions

What do you think of when you hear Machine Learning

Why Machine Learning Clustering Algorithms?

It is Critical to Have Good Data



What information is in your EHR?

Have you ever pulled aggregate data from your EHR?

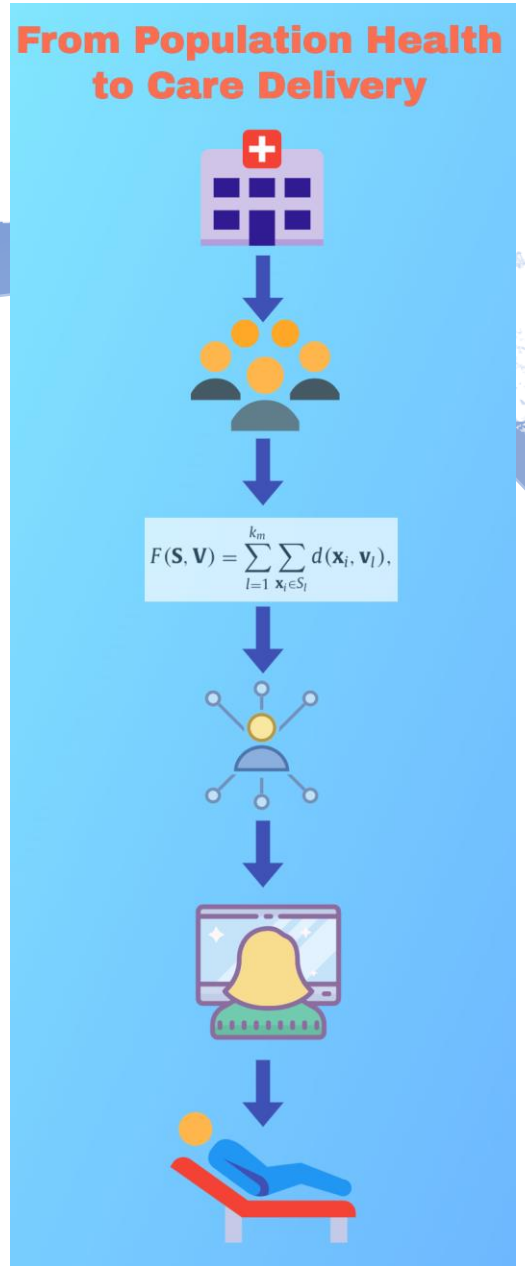
There may be many examples that could serve as good data

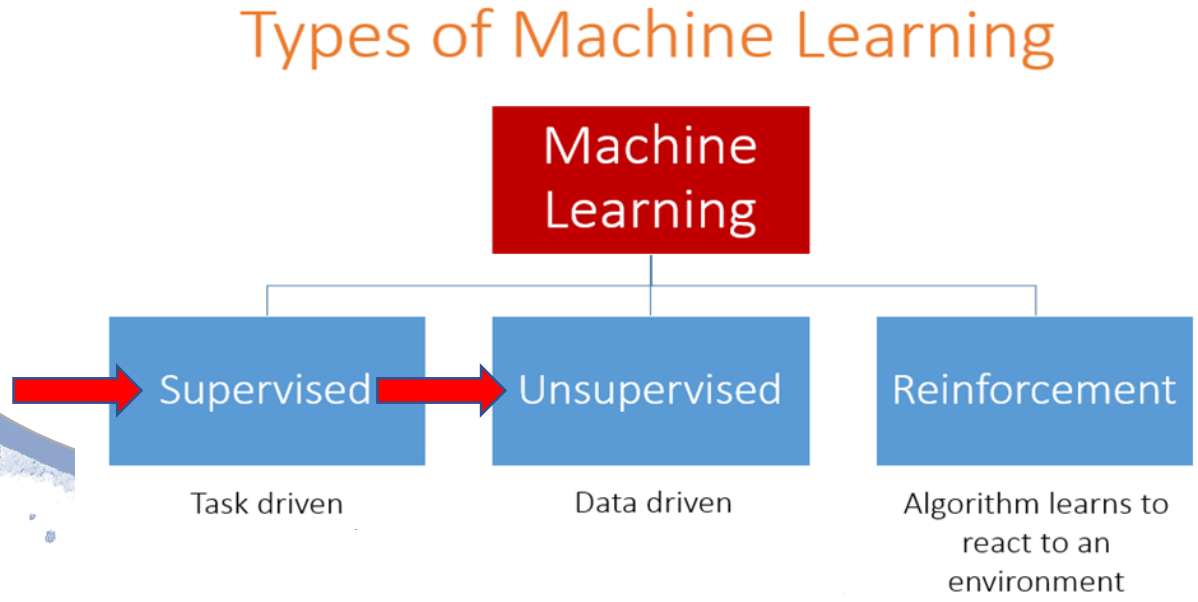
What might these be?

Examples: BIO-Psycho-Social-Spiritual Markers

Key Questions and Topics to Address with Clustering

- Health informatics- A different pathway to answering research questions.
- Identification of patient subgroups (i.e., which patients are most similar or dissimilar from one another)
- How do you get answers to issues that make a difference in your context:
  - Value Based Care
  - Cost efficiency
  - Frequent Use

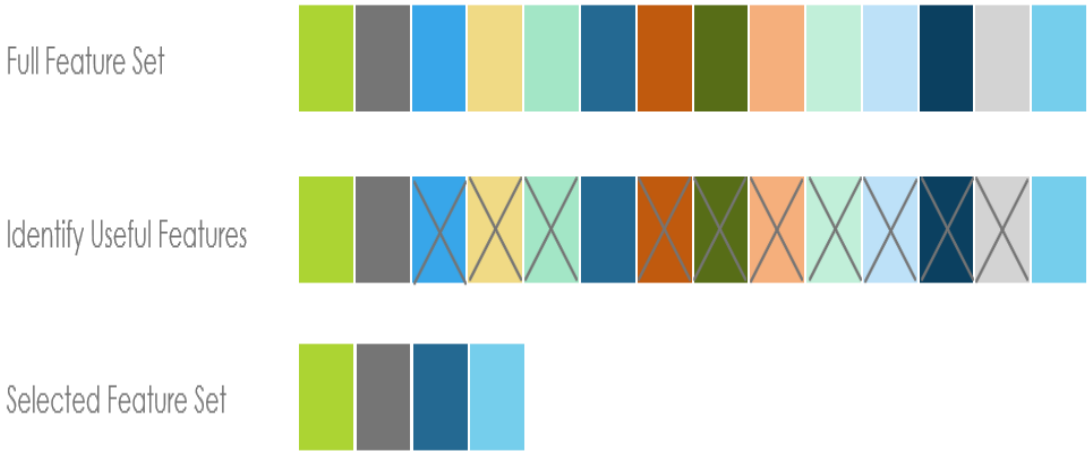




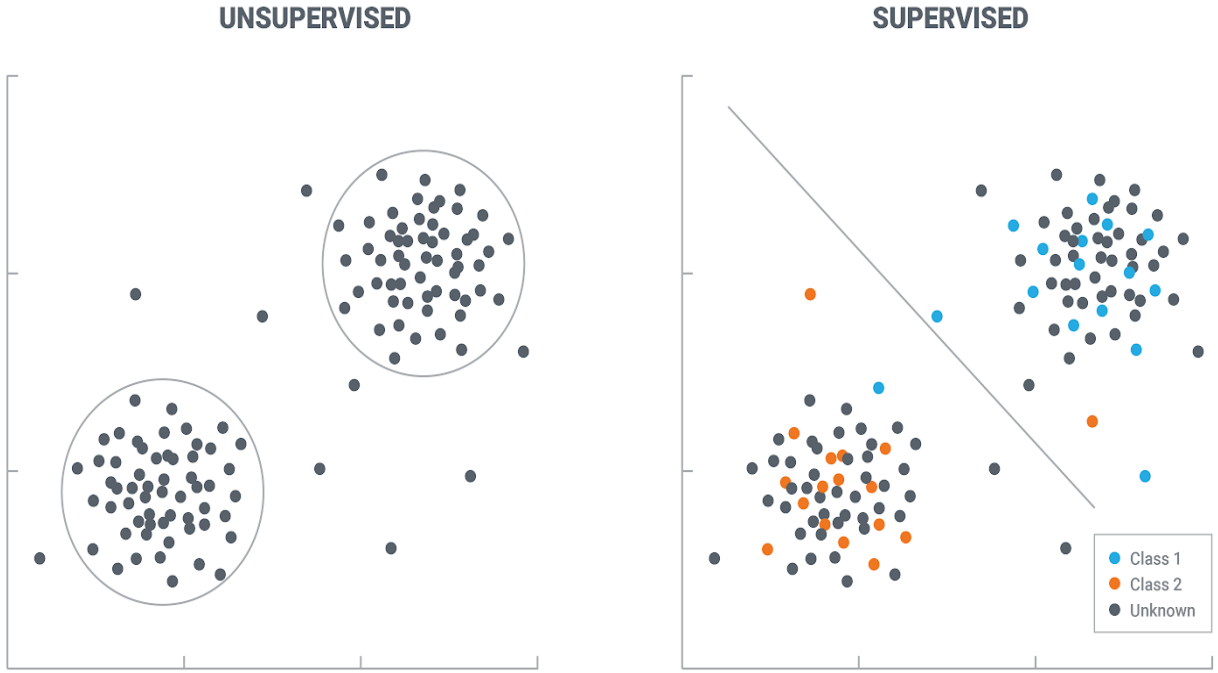
What is Machine Learning and How Can You Use It to Support Healthcare Delivery?

Clustering

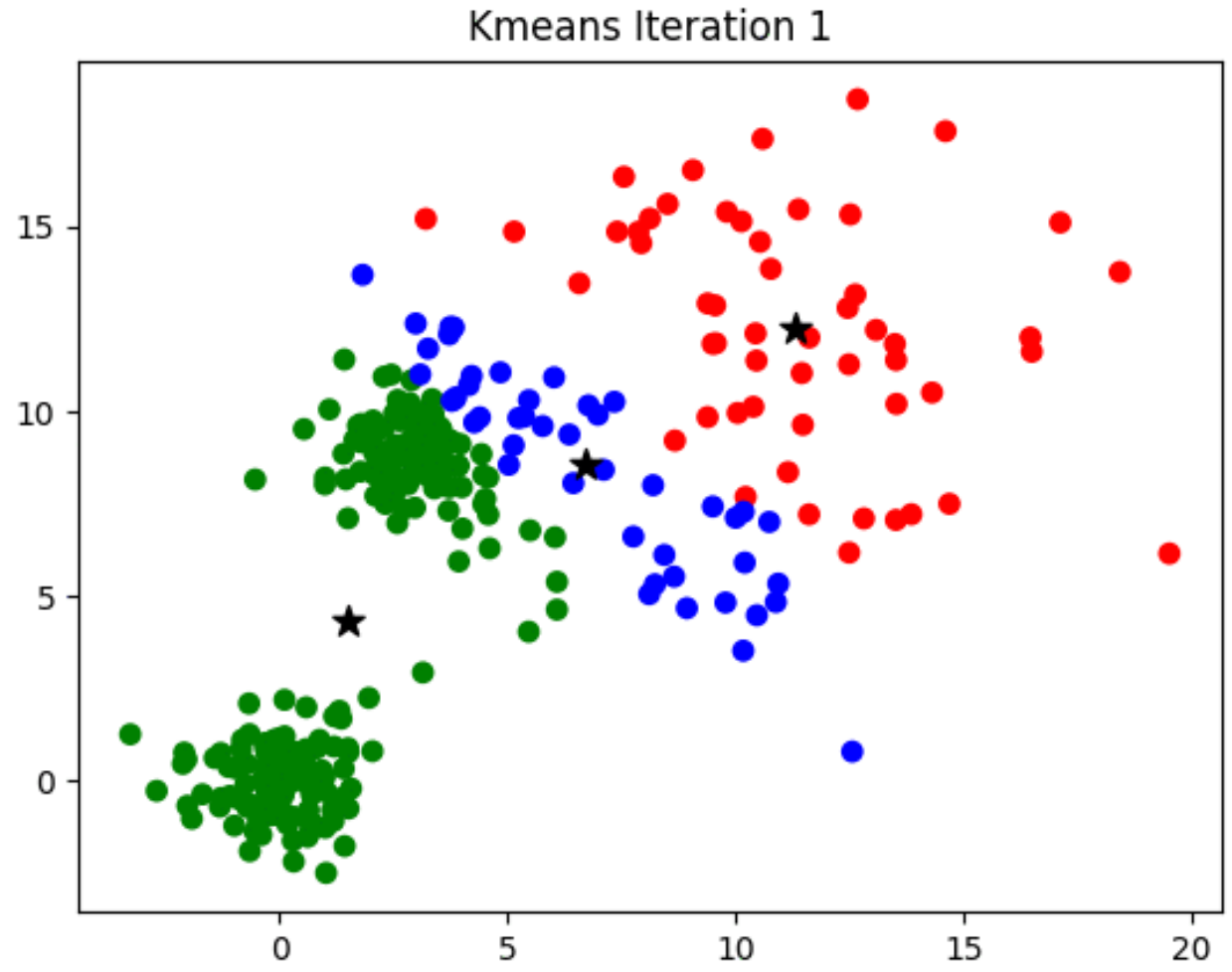
Feature Selection



Clustering



Clustering Algorithm Example: K-means k-3



Amelia, A. (2018).*K-means clustering: From A to Z.* Retrieved from https://towardsdatascience.com/k-means-clustering-from-a-to-z-f6242a314e9a

Results of Supervised Clustering

Supervised Clustering Algorithm Performance, Cohen's Kappa

Cluster Type	2011 Kappa	2012 Kappa	2013 Kappa	2014 Kappa	2015 Kappa	Avg Kappa/Performance
DBSCAN	0.526	0.529	0.534	0.548	0.549	0.5372
K-means K-3	0.54	0.523	0.517	0.556	0.591	0.5454
K-means K-4	0.521	0.519	0.536	0.531	0.539	0.5292
K-means K-5	0.57	0.565	0.523	0.565	0.585	0.5616
K-means K-6	0.519	0.559	0.562	0.57	0.553	0.5526
K-means K-7	0.548	0.554	0.554	0.551	0.535	0.5484
Support Vector	0.534	0.537	0.548	0.531	0.547	0.5394
X-means	0.525	0.562	0.543	0.574	0.553	0.5514
Topdown k-3	0.538	0.525	0.518	0.547	0.567	0.539
Topdown k-4	0.524	0.543	0.536	0.552	0.543	0.5396
Topdown k-5	0.54	0.54	0.523	0.546	0.552	0.5402
Topdown k-6	0.521	0.543	0.562	0.555	0.553	0.5468
Topdown k-7	0.514	0.53	0.55	0.572	0.562	0.5456
IQR	0.019	0.02525	0.029	0.021	0.01575	0.022
Lower Quartile	0.521	0.526	0.523	0.5465	0.544	0.5321
Upper Quartile	0.54	0.55125	0.552	0.5675	0.55975	0.5541

Cohen's Kappa = 0.01–0.20 as none to slight, 0.21–0.40 as fair, 0.41– 0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as almost perfect agreement

Recall and Precision Scores for Supervised K-Means k-6 Algorithm

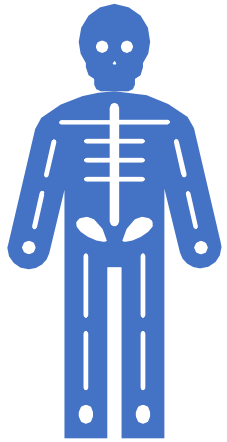
Year	1-2 Visits Recall	1-2 Visits Precision	3+ Visits Recall	3+ Visits Precision	4+ Visits Recall	4+ Visits Precision	5+ Visits Recall	5+ Visits Precision	6+ Visits Recall	6+ Visits Precision
2011	99.29	93.6	23.55	40.35	26.71	36.63	16	32.88	66.81	92.44
2012	99.32	94.45	38.36	43.6	17.52	28.95	14.63	33.8	57.66	96.93
2013	99.3	93.88	30.78	44.38	33.12	40.8	8.97	35.9	69.42	95.54
2014	99.29	94.27	35.89	43.47	23.28	36.79	12.28	40.38	68.56	94.91
2015	99.37	93.36	27.89	43.62	27.87	37.16	16.85	44.29	71.13	97.15

Results of Unsupervised Clustering

Unsupervised Clustering Algorithm Performance

Cluster Type	2011 G	2011 SSa	2012 G	2012 SSa	2013 G	2013 SSa	2014 G	2014 SSa	2015 G	2015 SSa
DBSCAN	0.997	0.966	0.997	0.966	0.997	0.969	0.997	0.963	0.997	0.967
K-means K-3	1	0.65	1	0.64	1	0.636	1	0.627	1	0.622
K-means K-4	1	0.62	1	0.602	1	0.614	1	0.591	1	0.587
K-means K-5	1	0.59	1	0.588	1	0.596	1	0.576	1	0.577
K-means K-6	1	0.579	1	0.587	1	0.591	1	0.575	1	0.566
K-means K-7	1	0.578	1	0.583	1	0.588	1	0.572	1	0.562
Support Vector	1	1	1	1	1	1	1	1	1	1
X-means	1	0.65	1	0.638	1	0.636	1	0.627	1	0.624
Topdown k-3	1	0.61	1	0.605	1	0.597	1	0.591	1	0.589
Topdown k-4	1	0.33	1	0.326	1	0.321	1	0.318	1	0.316
Topdown k-5	1	0.312	1	0.309	1	0.305	1	0.302	1	0.301
Topdown k-6	1	0.224	1	0.224	1	0.226	1	0.226	1	0.228
Topdown k-7	1	0.175	1	0.174	1	0.174	1	0.217	1	0.219
IQR		0.329		0.3215		0.323		0.317		0.3145
Lower Quartile		0.321		0.3175		0.313		0.31		0.3085
Upper Quartile		0.65		0.639		0.636		0.627		0.623

Number of ED Visits Is Important For Grouping Patients, BUT…..
Only As Part Of A Complete, Systemic Picture



- **Variables that distinguish the patient clusters from one another across years:**

- Payer type (i.e., Medicare, Medicaid, private insurance, self-pay, no charge, other)
- Select diagnostic categories (i.e., mental illness, nervous system and sense organ disorders)
- Number of ED visits per year

How to Make Meaning of Clustering Results

Rapidminer

RapidMiner is a data science software platform that provides an integrated environment for data preparation, machine learning, deep learning, text mining, and predictive analytics

Now let’s take a look!

How Could You Apply This At Your Own Site?

Jessica Goodman, PhD
Jessica_Goodman@URMC.Rochester.edu

Angela Lamson, PhD, LMFT
lamsona@ecu.edu

Questions?

Video Tutorials

- https://www.youtube.com/watch?v=lZho66YQEIM&t=9s
- https://www.youtube.com/watch?v=lZho66YQEIM
- https://www.youtube.com/watch?v=E-el-z06-g8
- https://www.youtube.com/watch?v=G0AM6-KtauY

Session Survey

Use the CFHA mobile app to complete the survey/evaluation for this session.

CFHA COLLABORATIVE FAMILY HEALTHCARE ASSOCIATION

Join us next year in Philadelphia, Pennsylvania! Thank you!