Choosing your machine learning algorithm

Is the data labeled?

Data contains ground truth about patients' disease status

Yes

Make predictions

Continuous outcomes

Ex: Cost of care in $

Regression

- Linear Model
  - Lasso
  - Ridge
- Random Forest
- k-Nearest Neighbors
- Linear Mixed Model
- XGBoost
- Neural Network

Discrete outcomes

Ex: Diabetes Y/N

Classification

Listed by increasing sophistication

- Linear Model
  - Lasso
  - Ridge
- Random Forest
- k-Nearest Neighbors
- Support Vector Machine
- Linear Mixed Model
- XGBoost
- Neural Network

No

Explore variation

Pick a measure

Ex: Cost of care in $

Compare within categories

- Find Variation

Visualize difference between groups

- Variation Across Groups

Find groups

Need to group the data?

Clustering

- K-means

Need to reduce dimensions?

Dimension Reduction

- Principle Component Analysis

Available in healthcareai-r

Available in healthcareai-py

* Coming soon