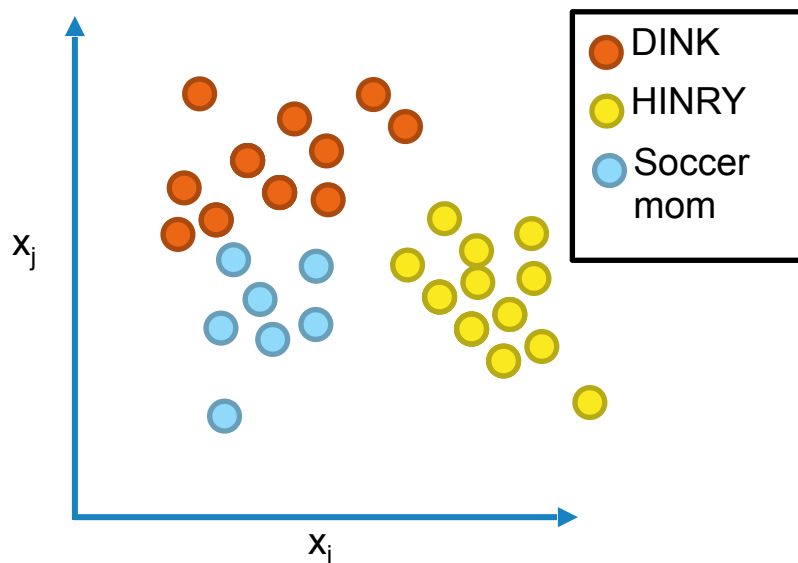


Clustering:

Grouping rows – e.g. creating groups of similar customers

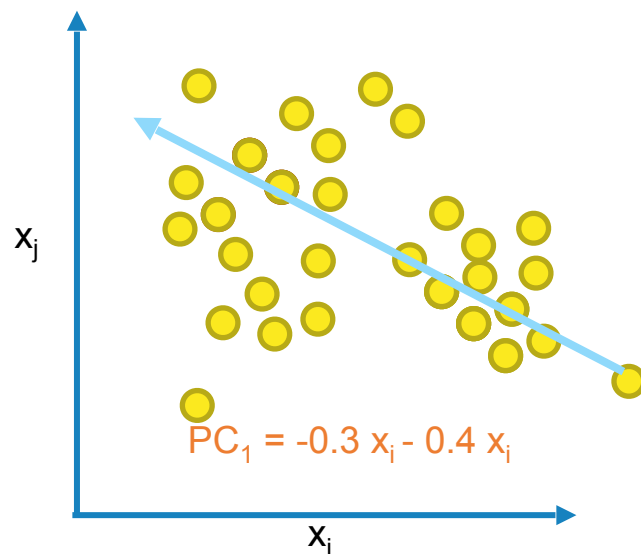


H₂O algos:

k – means

Feature extraction:

Grouping columns – Create a small number of new representative dimensions

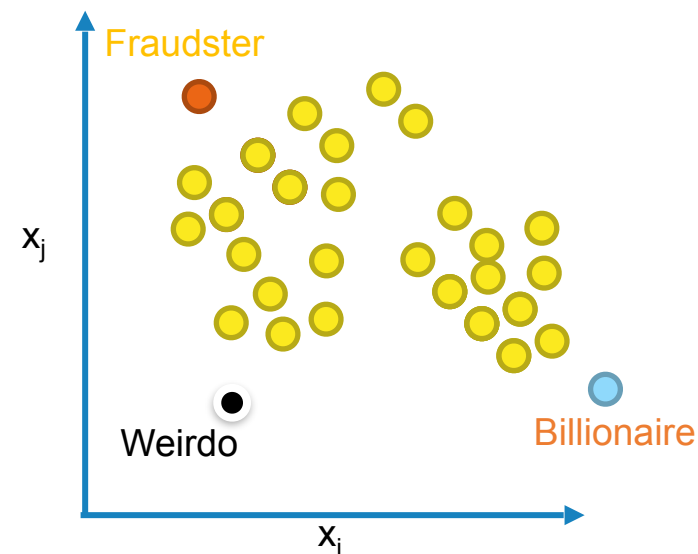


H₂O algos:

Principal components
Generalized low rank models
Autoencoders
Word2Vec

Anomaly detection:

Detecting outlying rows - Finding high-value, fraudulent, or weird customers



H₂O algos:

Principal components
Generalized low rank models
Autoencoders

Unsupervised Learning

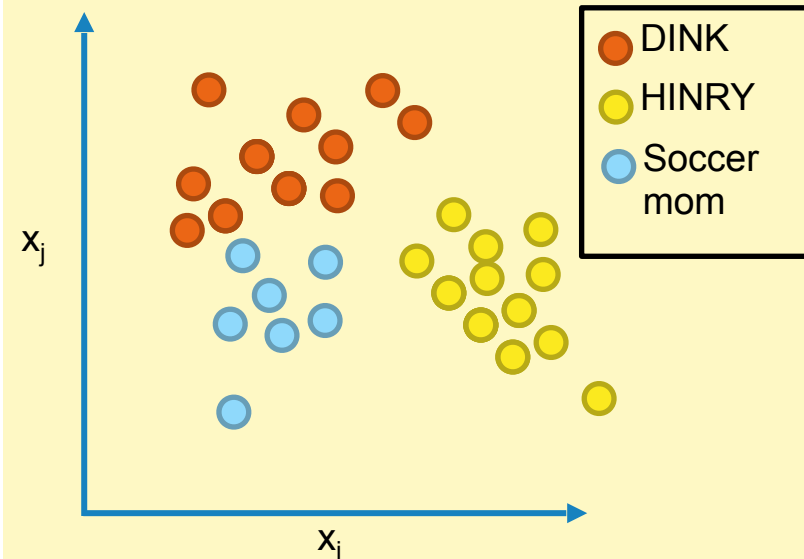
Vocabulary

Concept	Statistics\Econometrics	Machine Learning
“Computation”	Fit\Estimate	Train
“Left-hand side”	Dependent variable	Target
“Right-hand side”	Regressor\Predictor\Class	Feature\Factor\Enum
“Goal”	Estimation\Explanation	Prediction

Unsupervised Learning

Clustering:

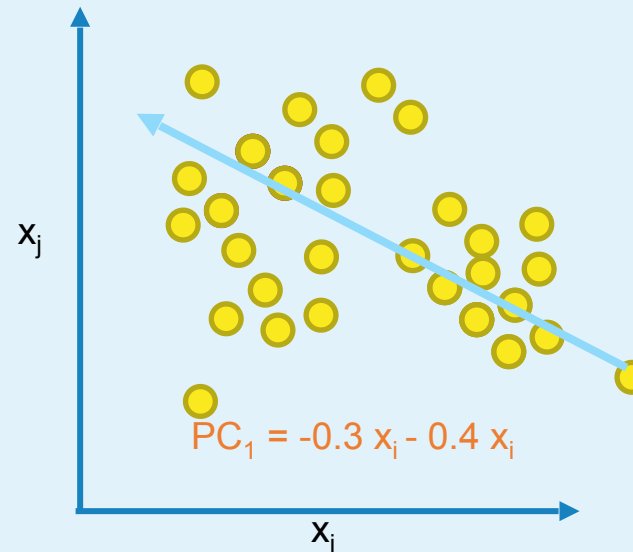
Grouping rows – e.g. creating groups of similar customers



H₂O algos:
k – means

Feature extraction:

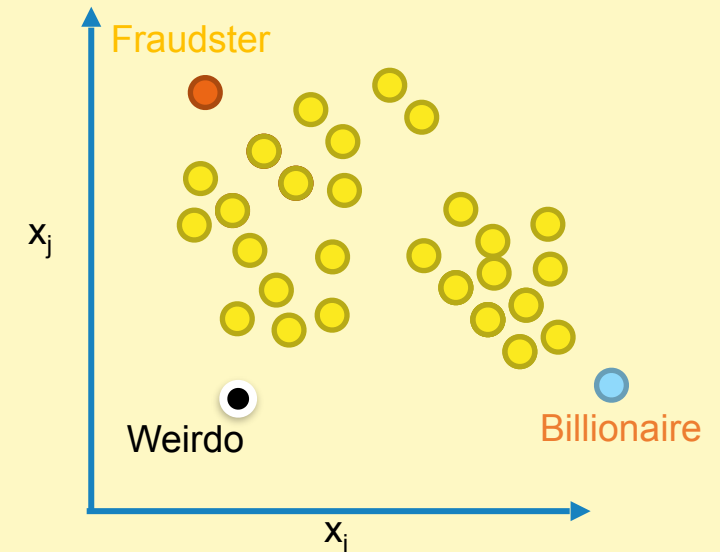
Grouping columns – Create a small number of new representative dimensions



H₂O algos:
Principal components
Generalized low rank models
Autoencoders
Word2Vec

Anomaly detection:

Detecting outlying rows - Finding high-value, fraudulent, or weird customers



H₂O algos:
Principal components
Generalized low rank models
Autoencoders