Ai+ TRAINING

Introduction to Anomaly Detection in Python & R

Dr. Aric LaBarr

Associate Professor of Analytics

www.ariclabarr.com

Course Outline

- Introduction
- Data Preparation
- Probability & Statistical Techniques
- Machine Learning Techniques
- Conclusion

Course Outline

- Introduction
    - Who am I?
    - What are Anomalies?
    - Anomaly Detection Analytical Framework
- Data Preparation
- Probability & Statistical Techniques
- Machine Learning Techniques
- Conclusion

Course Outline

- Introduction
- Data Preparation
    - Feature Engineering
    - Recency and Frequency
    - Categorical Feature Engineering
- Probability & Statistical Techniques
- Machine Learning Techniques
- Conclusion

Course Outline

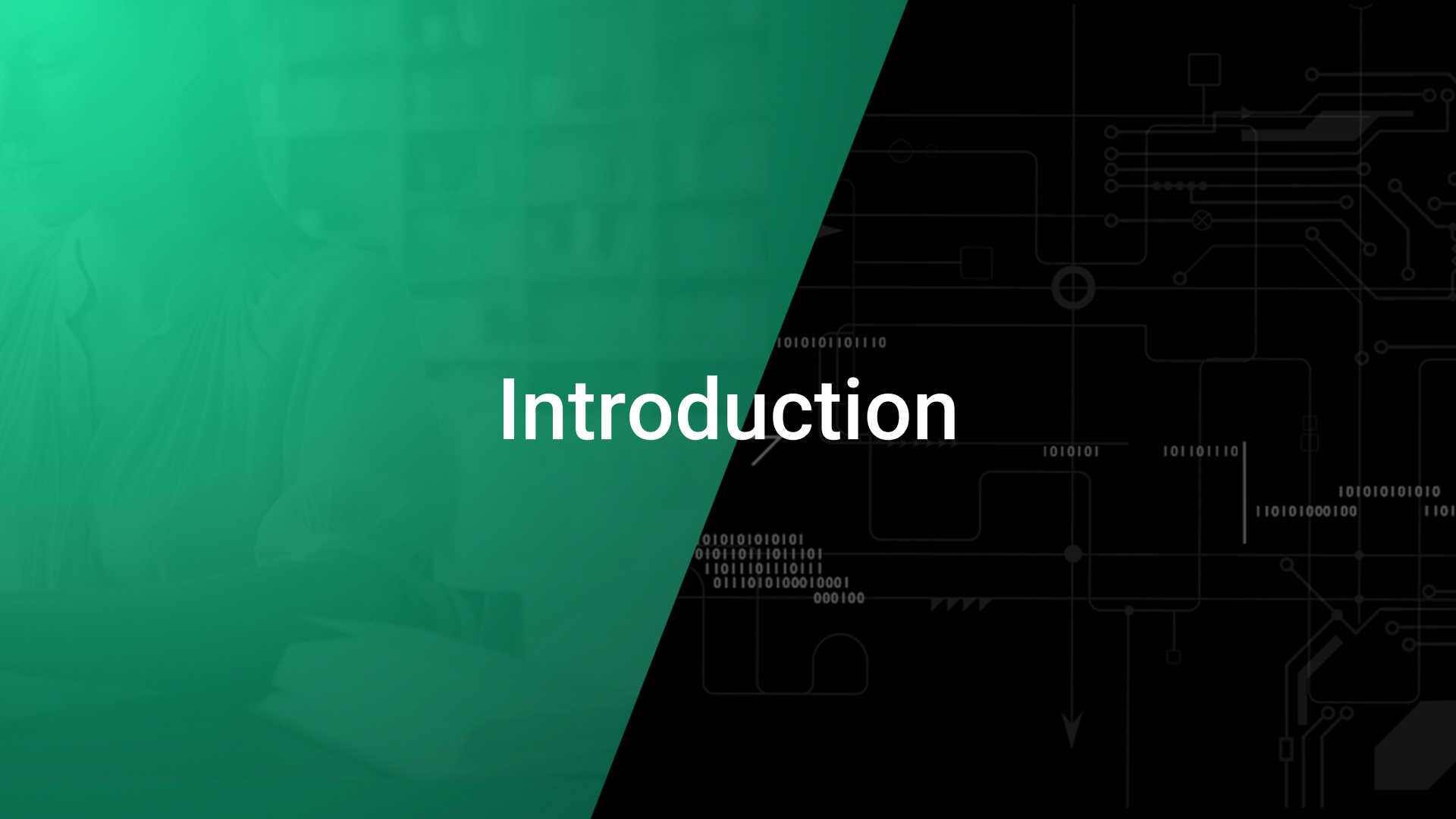
- Introduction
- Data Preparation
- Probability & Statistical Techniques
    - Benford’s Law
    - Z-scores and Robust Z-scores
    - IQR Rule and Its Adjustment
    - Mahalanobis Distances and Robust Mahalanobis
- Machine Learning Techniques
- Conclusion

Course Outline

- Introduction
- Data Preparation
- Probability & Statistical Techniques
- Machine Learning Techniques
    - k-Nearest Neighbors (k-NN)
    - Local Outlier Factor (LOF)
    - Isolation Forests
    - Classifier-Adjusted Density Estimation (CADE)
    - One-Class Support Vector Machine (SVM)
- Conclusion

Coding in Action

Example


Introduction

Introduction

- Introduction
  - Who am I?
  - What are Anomalies?
  - Anomaly Detection Analytical Framework

Introduction

Who Am I?

- Introduction
  - Who am I?
  - What are Anomalies?
  - Anomaly Detection Analytical Framework

Who Am I?

- 4-time North Carolina State University graduate:
  - BS in Statistics
  - BS in Economics
  - MS in Statistics
  - PhD in Statistics with minor in Economics

Who Am I?

- 4-time North Carolina State University graduate
- Former Senior Data Scientist and Director at Elder Research Inc.
    - Passionate about helping people solve challenges using their data.
    - Mentored a team of data scientists and software engineers to work closely with clients and partners to solve problems in predictive modeling, advanced analytics, forecasting, and risk management.

Who Am I?

- 4-time North Carolina State University graduate
- Former Senior Data Scientist and Director at Elder Research Inc.
- Associate Professor of Analytics at Institute for Advanced Analytics at NC State University
    - Nation's first master of science in analytics degree program
    - Helped design the innovative program to prepare a modern work force to wisely communicate and handle a data-driven future.
    - Developed and taught courses in statistics, mathematics, finance, risk management, and operations research.

Who Am I?

- 4-time North Carolina State University graduate
- Former Senior Data Scientist and Director at Elder Research Inc.
- Associate Professor of Analytics at Institute for Advanced Analytics at NC State University
- Find me online:
    - https://www.linkedin.com/in/ariclabarr/
    - https://www.youtube.com/c/AricLaBarr/
    - https://www.ariclabarr.com/

Introduction

What are Anomalies?

- Introduction
  - Who am I?
  - What are Anomalies?
  - Anomaly Detection Analytical Framework

What is an Anomaly?

anomaly

noun

/əˈnäməlē/

something that **deviates** from what is **standard**, **normal**, or **expected**

Why Detect Anomalies?

- Anomalies in data can lead to incorrect or out of date decisions to be made.
- Need to find these **outliers** before they become too much of a problem.
- Anomaly detection techniques used in variety of areas:
    - Cleaning data
    - Monitoring health of computer systems
    - Cybersecurity threats
    - Fraudulent claims or transactions

Why Detect Anomalies?

- Anomalies in data can lead to incorrect or out of date decisions to be made.
- Need to find these **outliers** before they become too much of a problem.
- Anomaly detection techniques used in variety of areas:
    - Cleaning data
    - Monitoring health of computer systems
    - Cybersecurity threats
    - Fraudulent claims or transactions

What is Fraud?

fraud

noun

/frôd/

Wrongful or criminal **deception** intended to result in financial or personal **gain**

Fraud Problem

- In 2022, the ACFE (Association of Certified Fraud Examiners) Report to the Nations estimated that organizations lose approximately 5% of their revenues to fraud.
- Based on 2022 estimated world GDP (IMF estimates) this would mean approximately $5.19 trillion is lost each year due to fraud.

Fraud Characteristics

1. Uncommon
2. Concealed and trying to be avoided
3. Ever changing and adapting
4. Thought out and organized
5. Doesn’t all look the same

Fraud Characteristics

1. Uncommon
2. Concealed and trying to be avoided
3. Ever changing and adapting
4. Thought out and organized
5. Doesn’t all look the same

- Because of these characteristics, fraud is a tough anomaly problem to solve.
- Data science can help aid in this problem!

Introduction

Anomaly Detection Analytical Framework

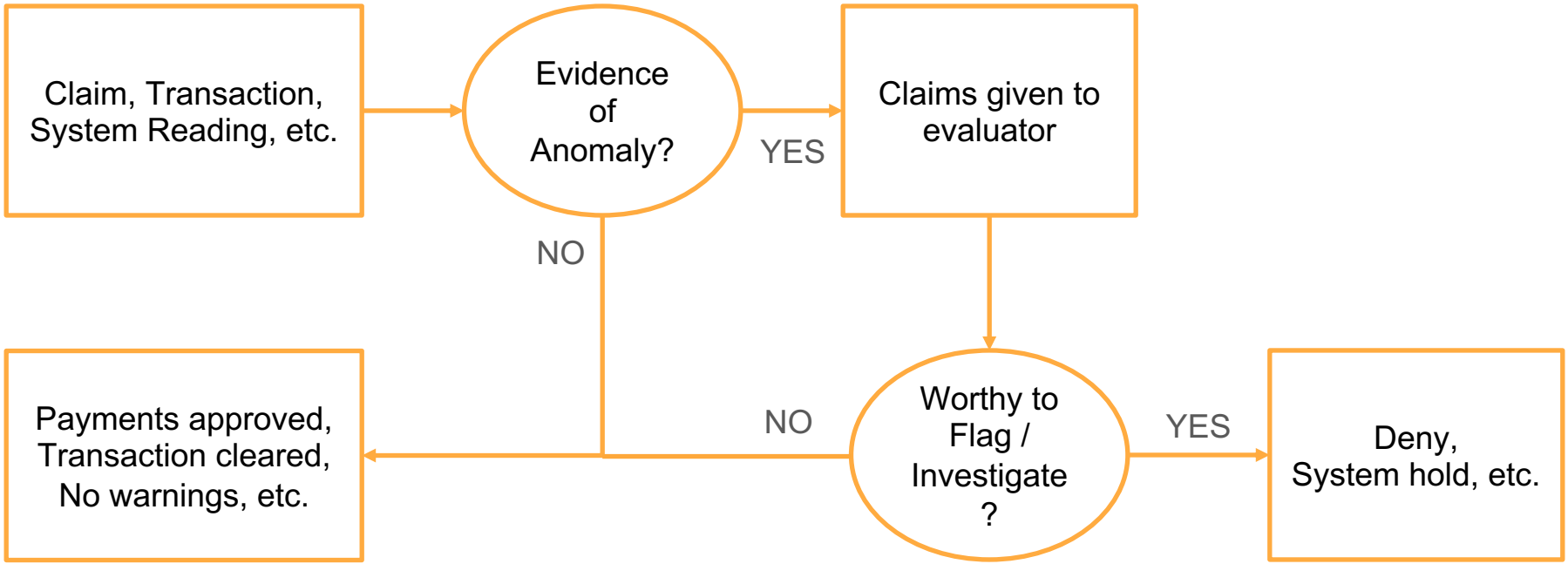
- Introduction
  - Who am I?
  - What are Anomalies?
  - Anomaly Detection Analytical Framework

Anomaly Detection Systems

- Regardless of the industry, two things are important for any anomaly detection solution or system:

    1. **DETECTION** – able to identify current anomalies in the system
    2. **PREVENTION** – able to flag potentially new anomalies in the system

Anomaly Detection Systems



Anomaly Detection Maturity – Card Transaction

- New / young anomaly detection solutions are based on **business rules**.
- Example:
  - IF:
    - Amount of transaction above threshold
  - THEN:
    - Flag as suspicious AND
    - Alert evaluator

Anomaly Detection Maturity – Cybersecurity

- New / young anomaly detection solutions are based on **business rules**.
- Example:
  - IF:
    - System pinged at unusual time
  - THEN:
    - Flag as suspicious AND
    - Alert evaluator

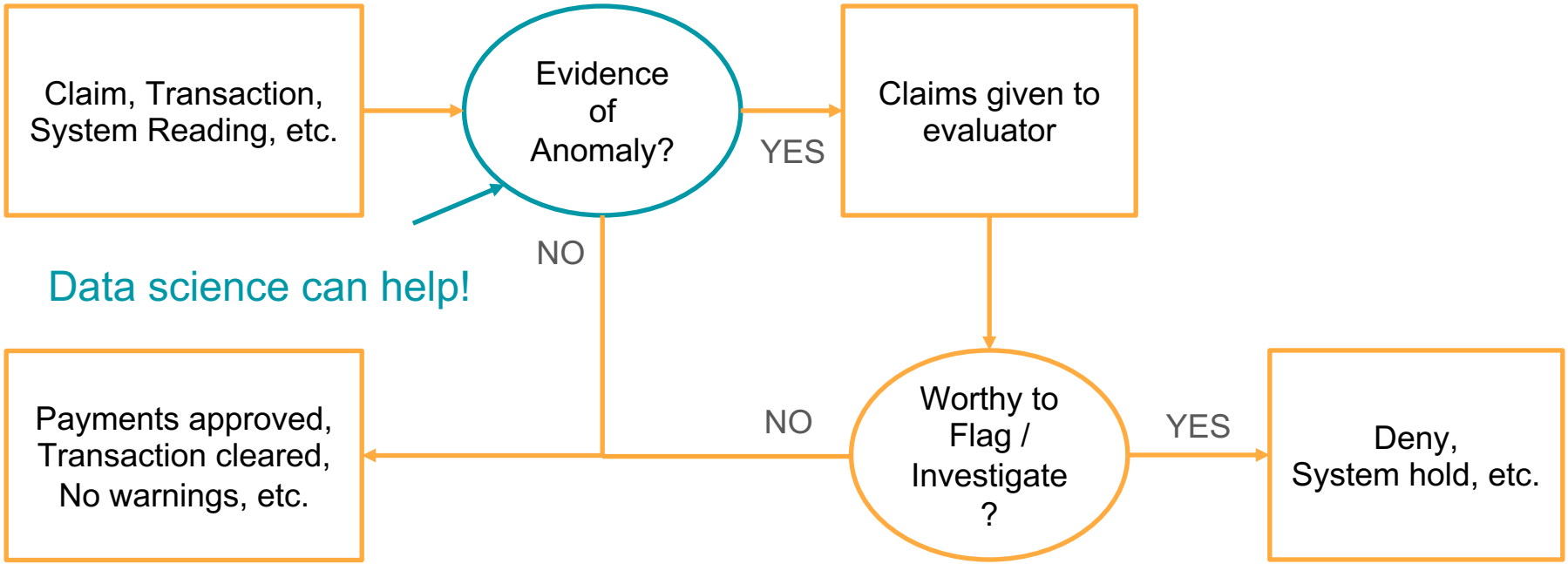
Anomaly Detection Maturity – Insurance Fraud

- New / young anomaly detection solutions are based on **business rules**.
- Example:
  - IF:
    - Severe injury but no doctor report
  - THEN:
    - Flag as suspicious AND
    - Alert evaluator

Business Rule Approach

- Advantages:
  - Simple
  - Easy to implement
- Disadvantages:
  - Expensive
  - Difficult to maintain and manage
  - Completely historical
  - Threats discover rules

Anomaly Detection Systems



Analytical Anomaly Detection Framework

- Advantages
    1. **Precision**
        - Increased detection power
        - More information used in decisions
        - More anomalies evaluated

Analytical Anomaly Detection Framework

- Advantages
    1. **Precision**
    2. **Efficiency in Operations**
        - Automated processing of claims
        - Ranked cases for evaluators

Analytical Anomaly Detection Framework

- Advantages
    1. **Precision**
    2. **Efficiency in Operations**
    3. **Efficiency in Costs**
        - Cheaper to long-run maintain
        - Quicker evaluation
        - Higher return on evaluations

Introduction

Conclusion

- Introduction
  - Who am I?
  - What are Anomalies?
  - Anomaly Detection Analytical Framework


Data Preparation

Data Preparation

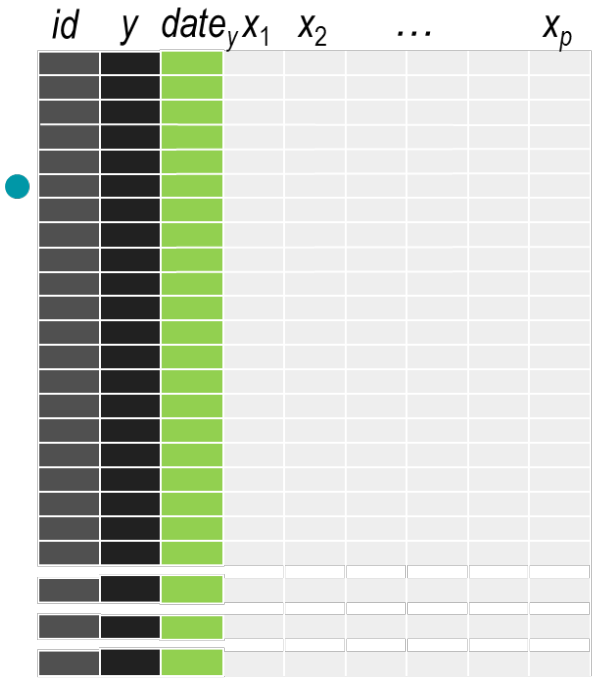
- Data Preparation
  - Feature Engineering
  - Recency and Frequency
  - Categorical Feature Engineering

Data Preparation

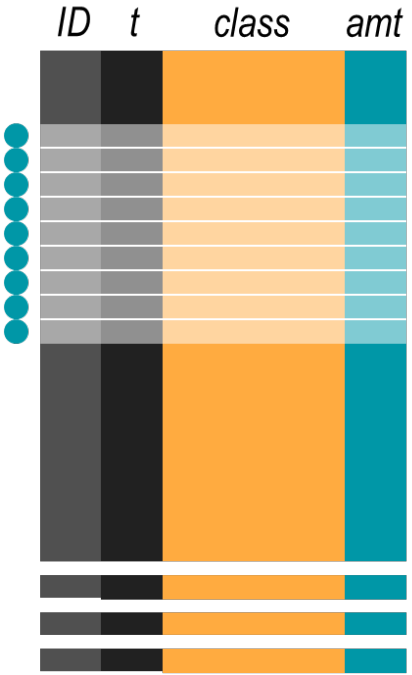
Feature Engineering

- Data Preparation
  - Feature Engineering
  - Recency and Frequency
  - Categorical Feature Engineering

Transaction Data



Model Development Data



Transaction Data

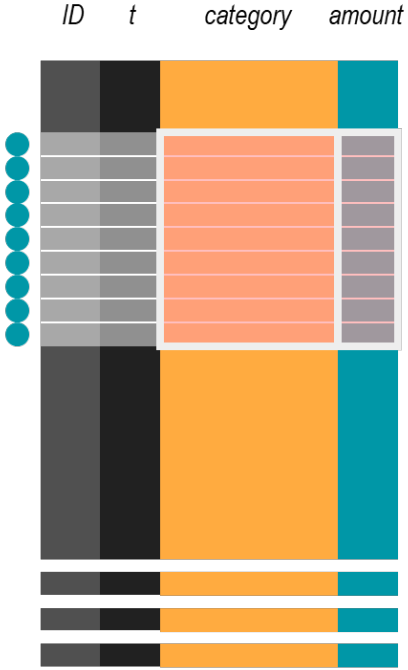
Transaction Data Examples

- There are many fields where transactional data plays an important role:
  - Credit card purchasing data
  - Medical / insurance claims data
  - Supply chain and logistics data
  - Censor / systems monitoring data
  - Etc.

Transactions Data

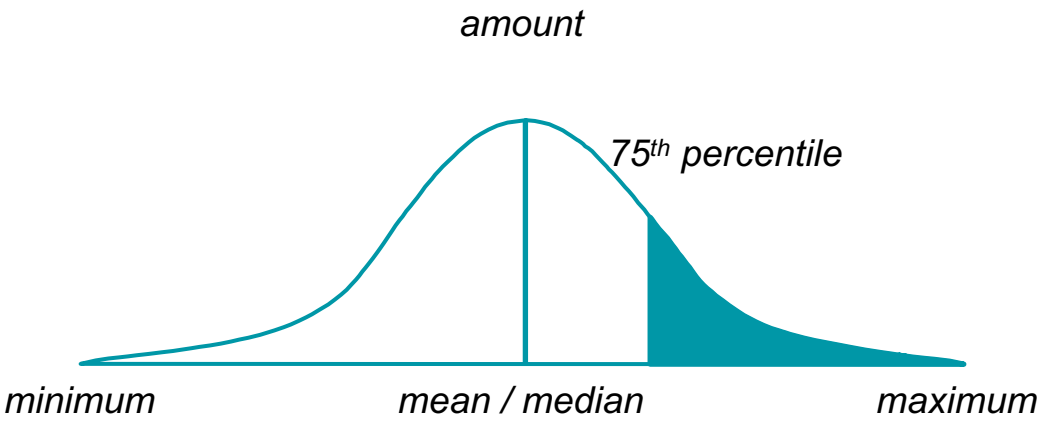
- Advantages
  - Highly detailed
  - Captures individual behavior
  - Strong prediction possible
- Challenges
  - Highly detailed
  - Difficult to obtain
  - Difficult to process

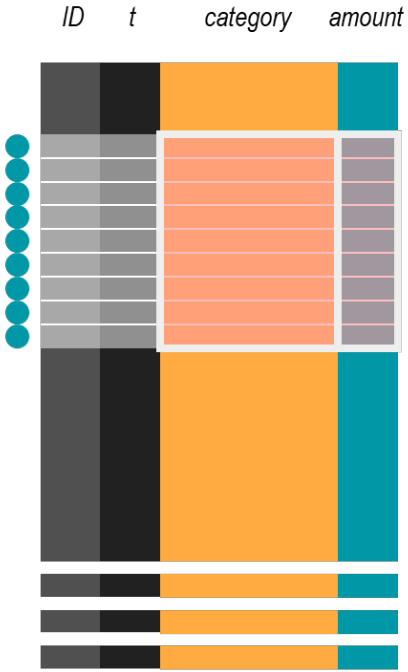
Input Possibilities: Tabulations



Transaction Data

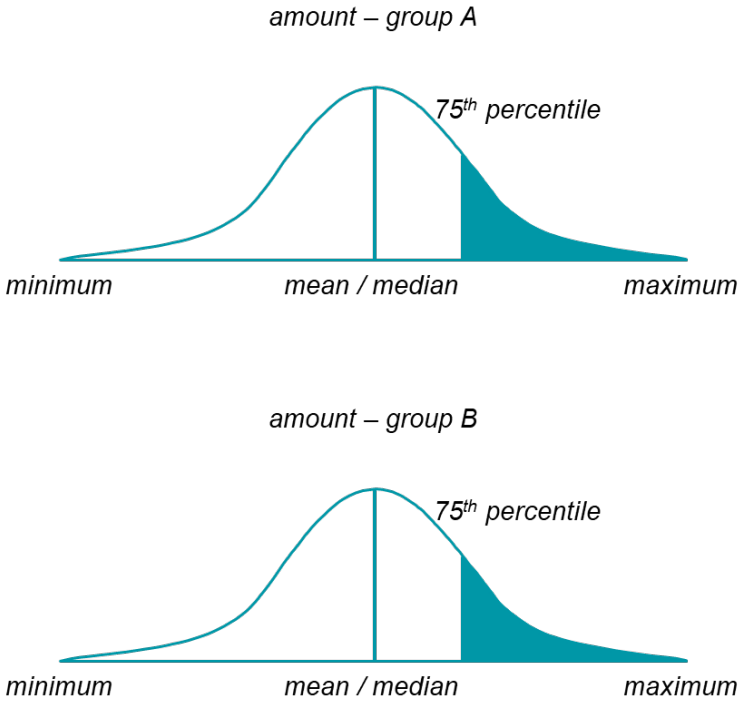
Input Possibilities: Tabulations

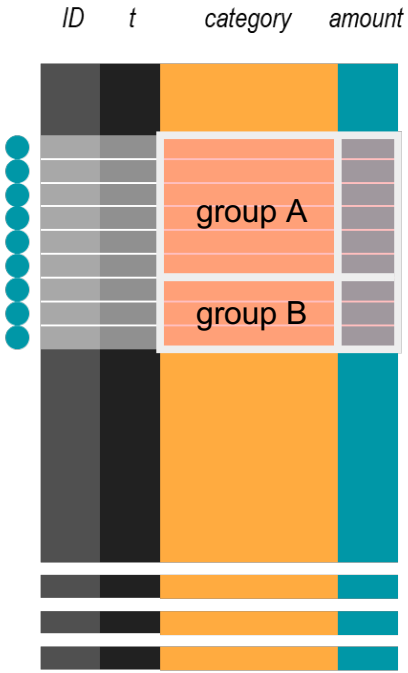




Transaction Data

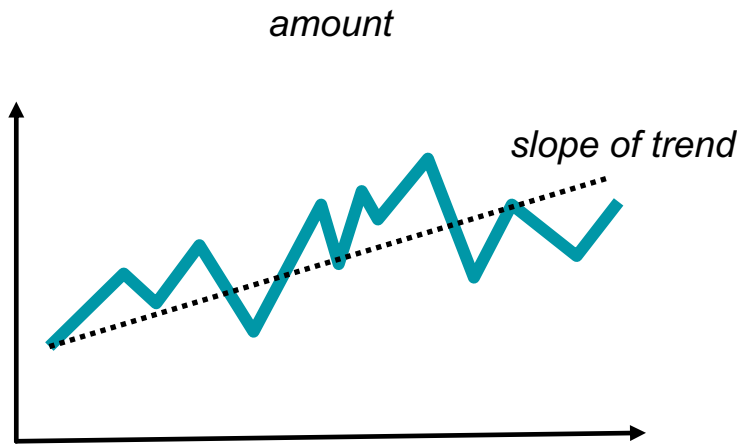
Input Possibilities: Tabulations

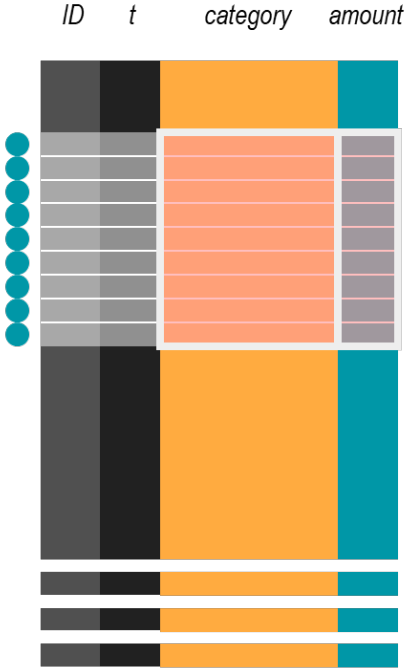




Transaction Data

Input Possibilities: Tabulations





Transaction Data

Coding in Action

Data Preparation – Feature Engineering

Data Preparation

Recency and Frequency

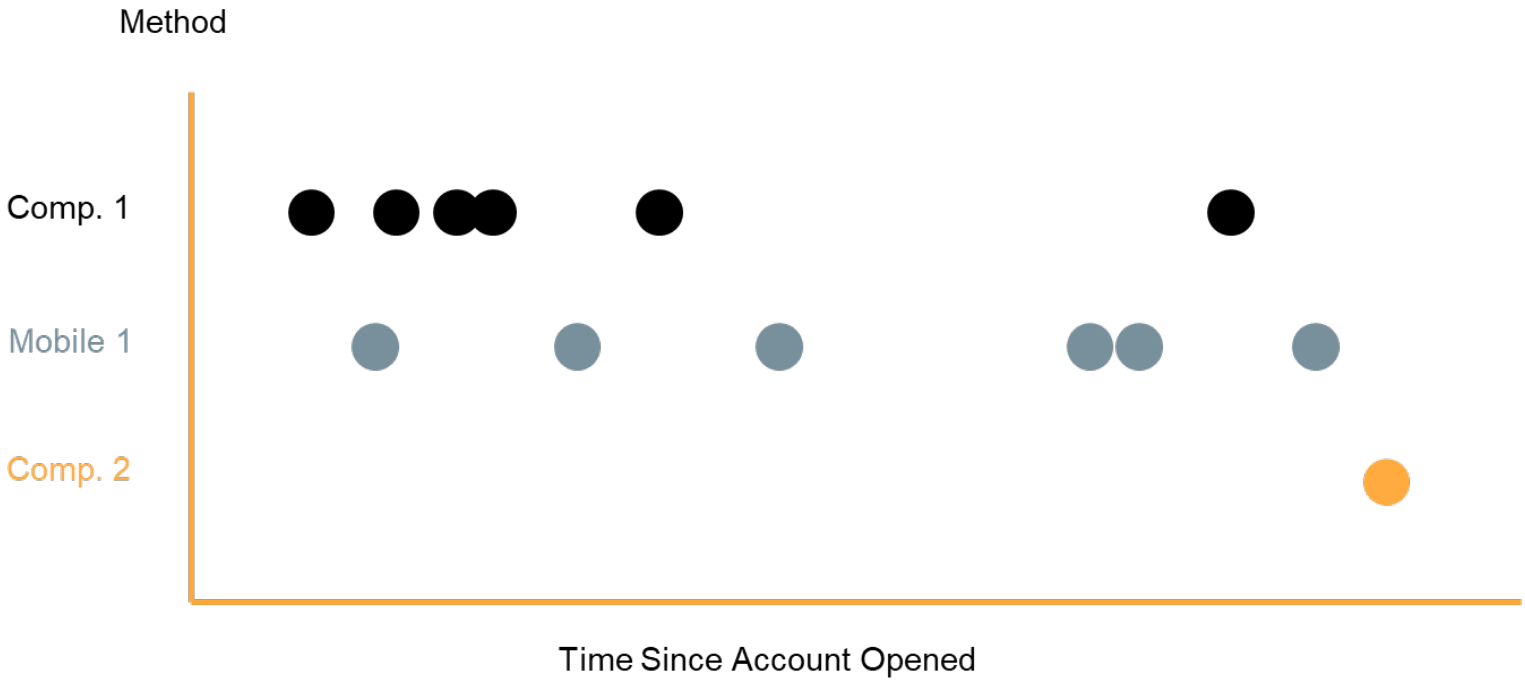
- Data Preparation
  - Feature Engineering
  - Recency and Frequency
  - Categorical Feature Engineering

Recency & Frequency

- Transactional data provides extensive information.
- Two of the most important things in fraud detection (as well as other fields) are **recency** and **frequency** of transaction.

- **Recency** – time in between transactions
- **Frequency** – how often transactions occur

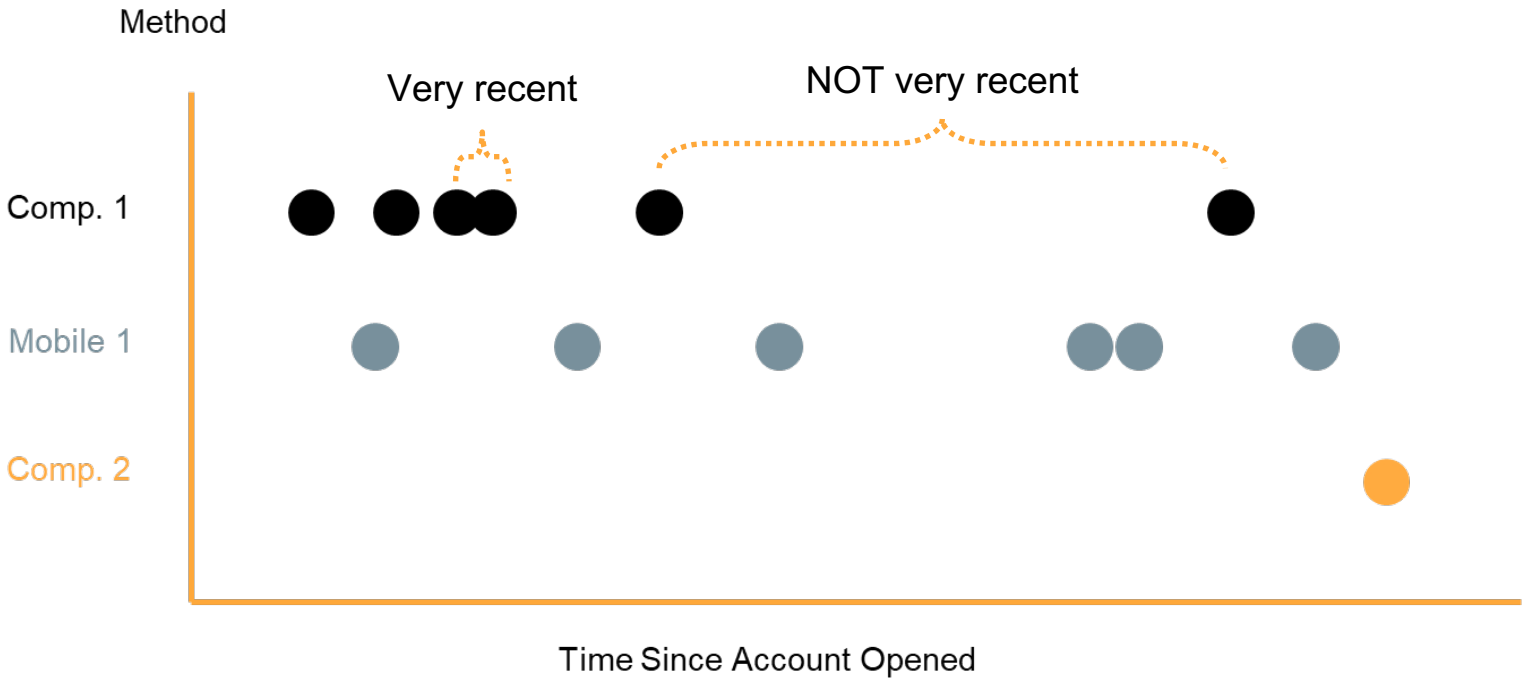
Online Account Access Example



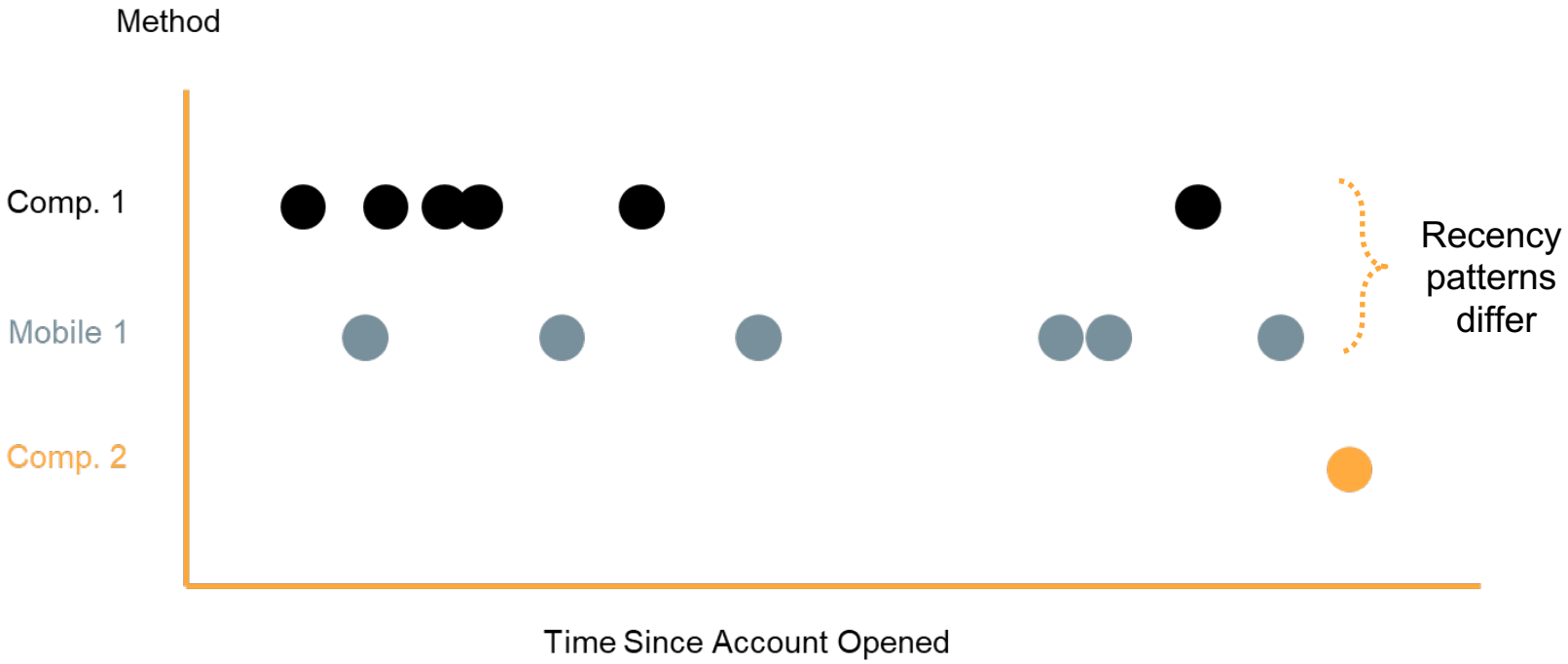
Recency

- **Recency** – time in between transactions
- Easy features:
  - Time in between transactions
  - Time since last transaction

Online Account Access Example



Online Account Access Example



Recency

- **Recency** – time in between transactions
- Easy features:
  - Time in between transactions
  - Time since last transaction
- What if you want to standardize these across different groups (authentication methods)?
- Is time symmetric?

Recency

- **Recency** – time in between transactions
- Easy features:
  - Time in between transactions
  - Time since last transaction
- What if you want to standardize these across different groups (authentication methods)?
- Is time symmetric? **NO – Careful how you standardize!**

Exponential Distribution

- Weight transactions based on their recency, but all on similar scale.
- **Exponential Distribution for Recency**:

$$Recency = e^{-\gamma t}$$

Exponential Distribution

- Weight transactions based on their recency, but all on similar scale.
- **Exponential Distribution for Recency**:

$$Recency = e^{-\gamma t}$$

Tuning parameter
for decline rate

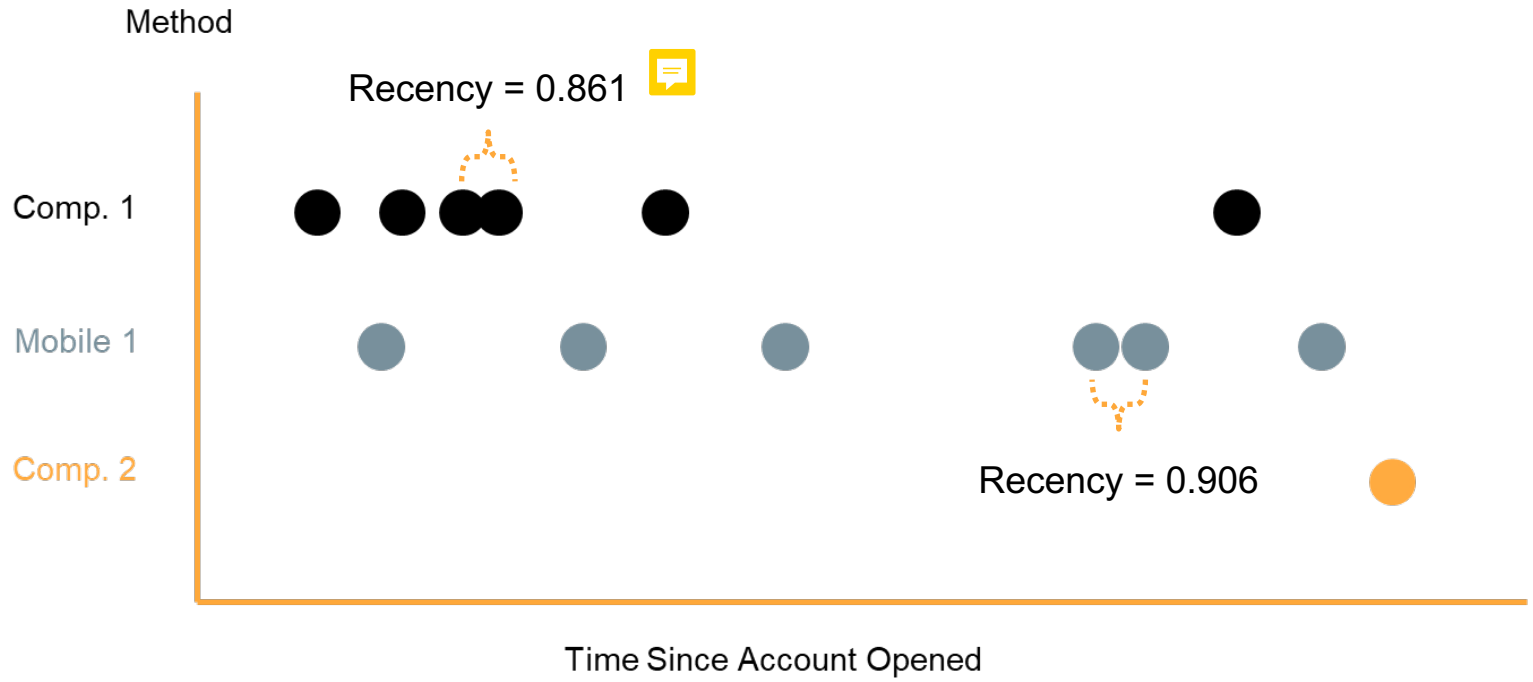
Exponential Distribution

- Weight transactions based on their recency, but all on similar scale.
- **Exponential Distribution for Recency**:

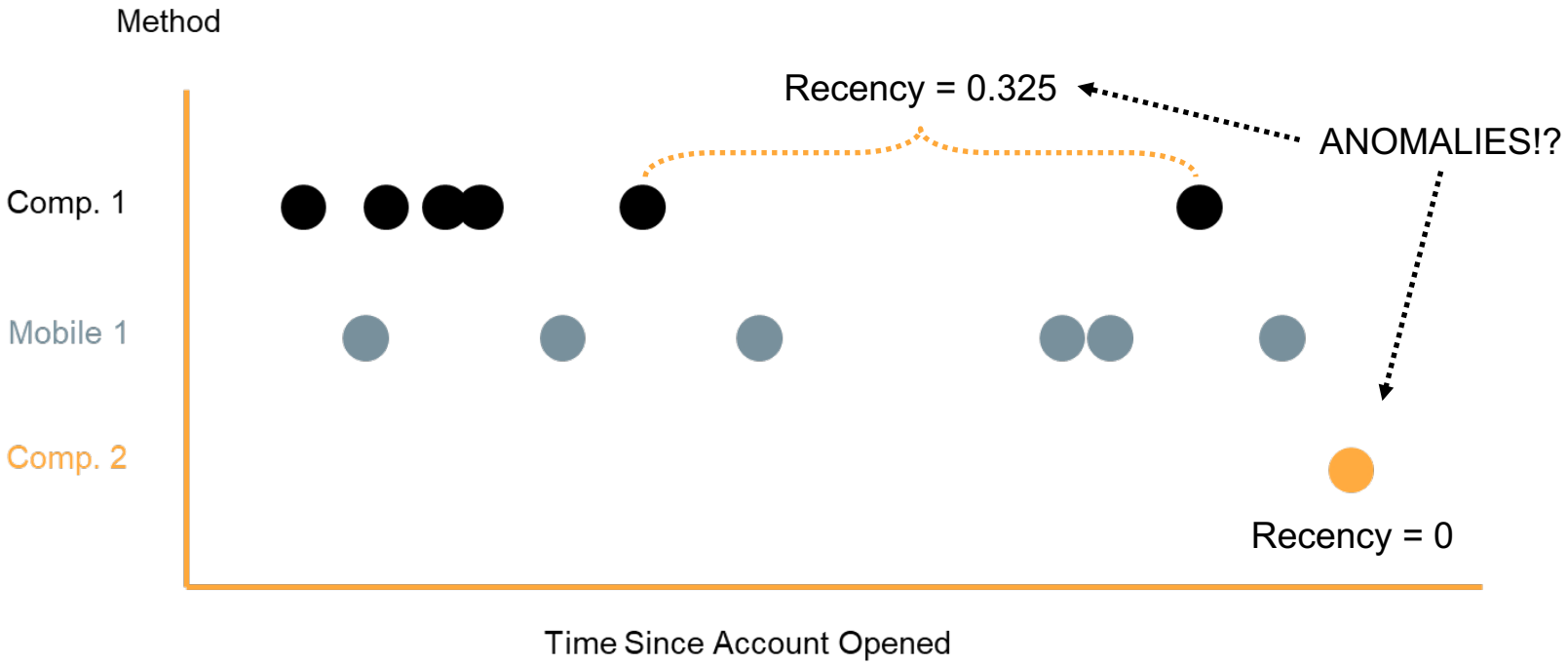
$$Recency = e^{-\gamma t}$$

- This forces recency to be bounded between 0 and 1.

Online Account Access Example



Online Account Access Example



How to Pick γ?

- Computer can optimize the parameter γ based solely on your data.
- Authentication Method:
    - Computer 1: $\gamma = -0.075$
    - Mobile 1: $\gamma = -0.0492$

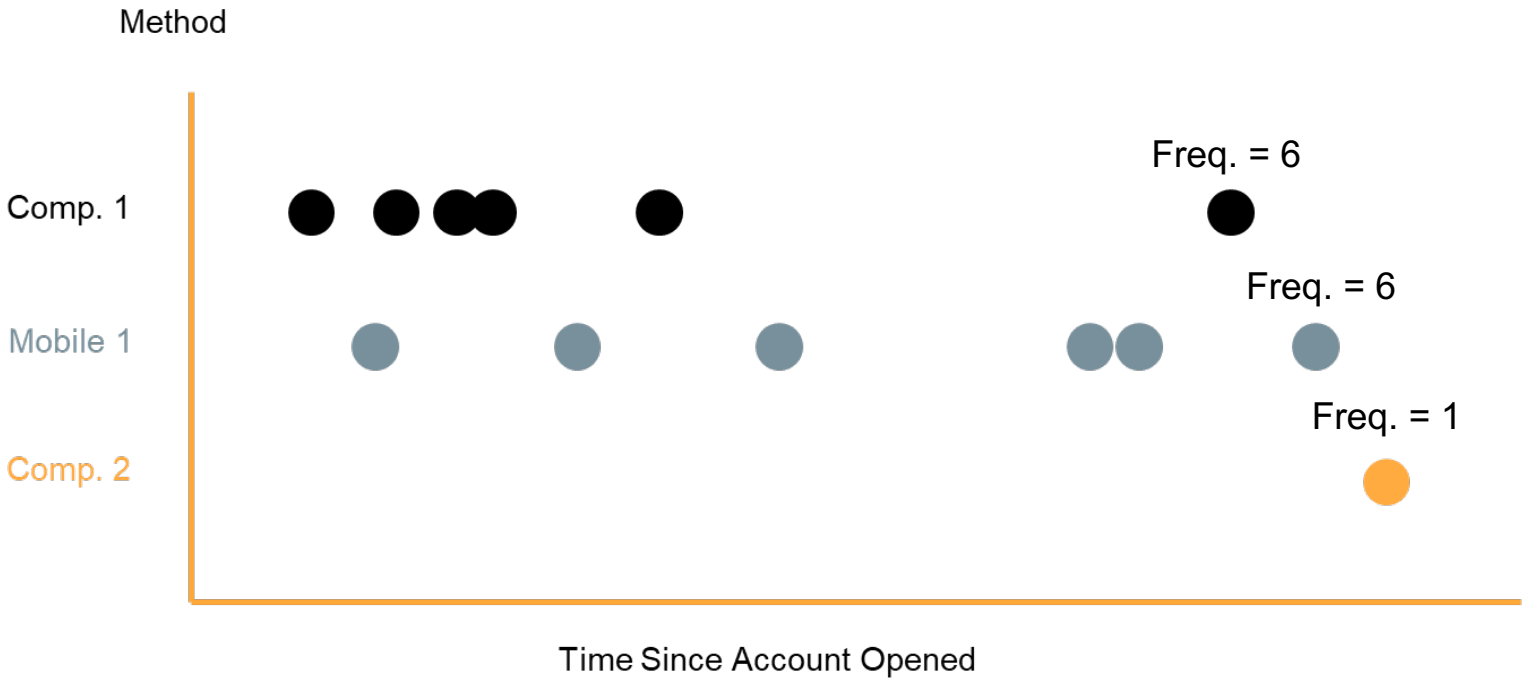
Coding in Action

Data Preparation – Recency

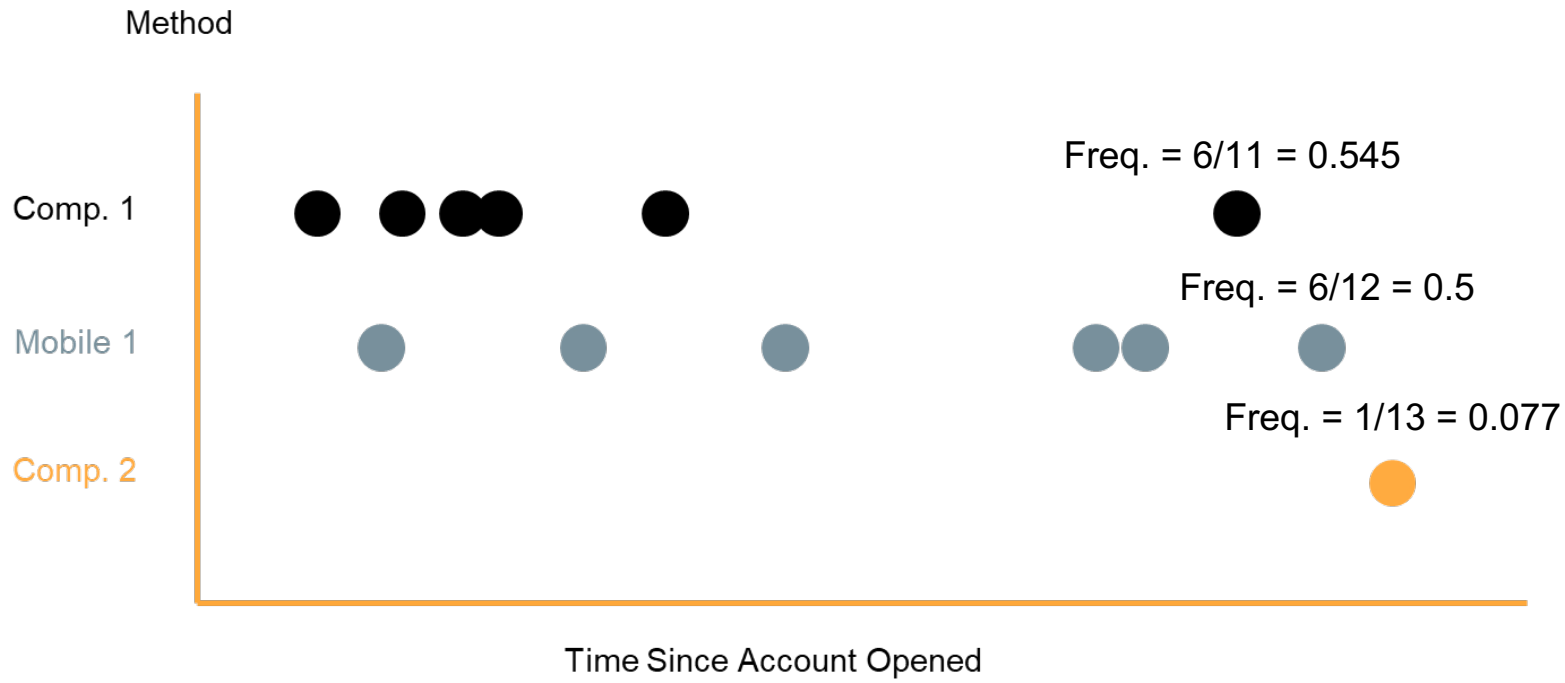
Frequency

- **Frequency** – how often transactions occur
- Easy features:
  - How many transactions total
  - How many transactions per group
  - Ratio of frequency by group to days active

Online Account Access Example



Online Account Access Example



Coding in Action

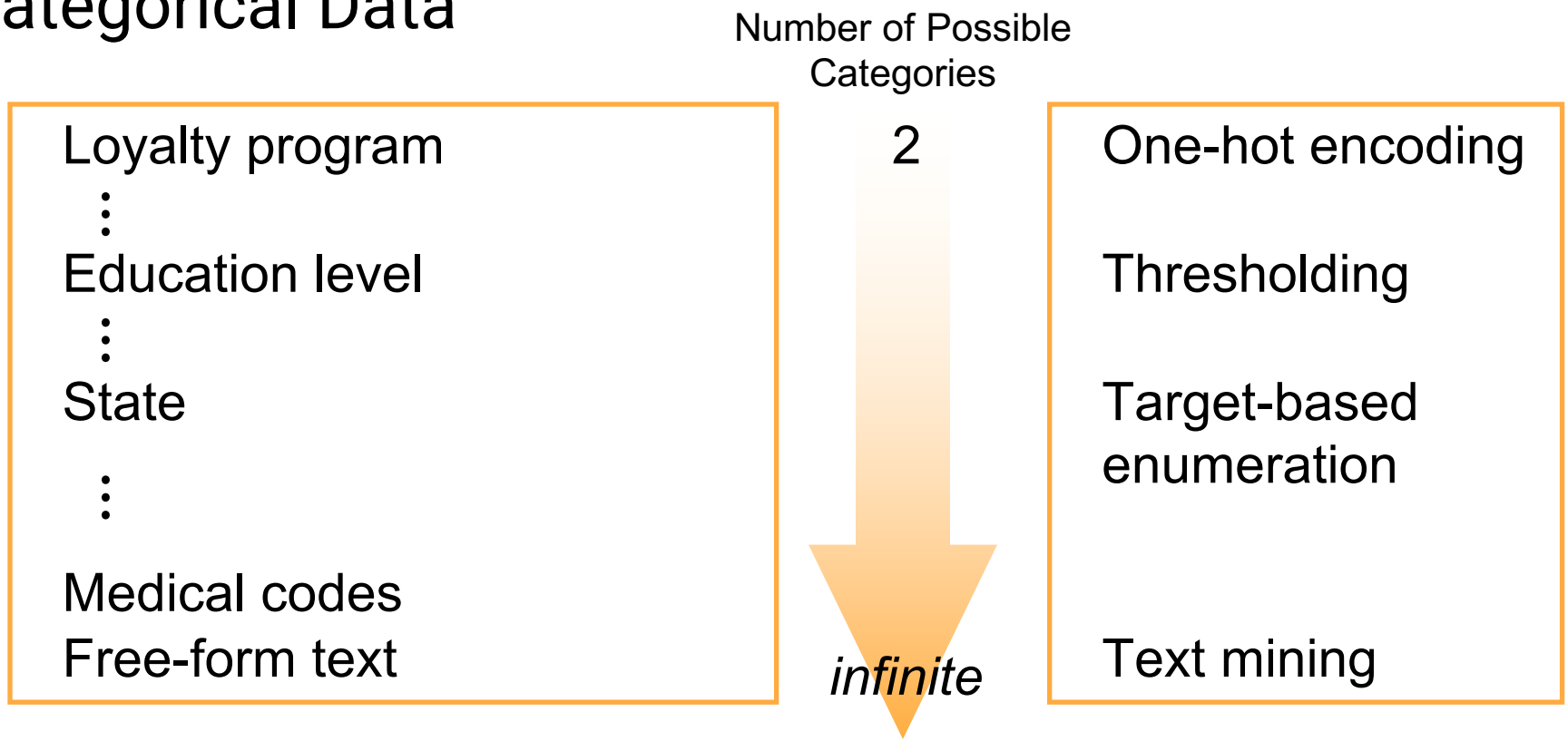
Data Preparation – Frequency

Data Preparation

Categorical Feature Engineering

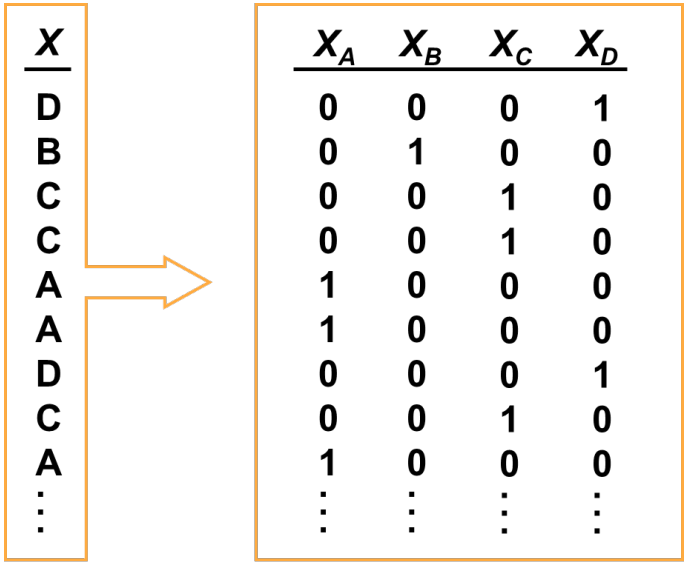
- Data Preparation
  - Feature Engineering
  - Recency and Frequency
  - Categorical Feature Engineering

Categorical Data



One-Hot Encoding

- With limited number of categories, it is easiest to just create binary representations of the categories in separate variables.



X
D
B
C
C
A
A
D
C
A
⋮

X_A	X_B	X_C	X_D
0	0	0	1
0	1	0	0
0	0	1	0
0	0	1	0
1	0	0	0
1	0	0	0
0	0	0	1
0	0	1	0
1	0	0	0
⋮	⋮	⋮	⋮

Thresholding

- When you have many categories, you can take the “important” categories (maybe ones with the largest counts) and let all the others be aggregated into an “other” category.

Level	Sample Size	Default
A	1562	332
B	970	53
C	223	17
D	111	10
E	85	4
F	50	10
G	23	1
H	17	0
I	12	1
J	5	1

Recombine into “Other” category

Target-Based Enumeration

- You can also represent categorical pieces of information numerically through target-based enumeration. Instead of the original categories, we now have a continuous variable for proportion of default.

Level	Sample Size	Default	Prop. of Default
A	1562	332	0.21
B	970	53	0.05
C	223	17	0.08
D	111	10	0.09
E	85	4	0.05
F	50	10	0.10
G	23	1	0.04
H	17	0	0.00
I	12	1	0.08
J	5	1	0.20

Level Clustering

- However, these new numeric representations can also be ordered and combined into new categories as well.

Level	Sample Size	Default	Prop. of Default
A	1562	332	0.21
B	970	53	0.05
C	223	17	0.08
D	111	10	0.09
E	85	4	0.05
F	50	10	0.10
G	23	1	0.04
H	17	0	0.00
I	12	1	0.08
J	5	1	0.20

Level Clustering

- However, these new numeric representations can also be ordered and combined into new categories as well.

Level	Sample Size	Default	Prop. of Default	New Level
A	1562	332	0.21	A
J	5	1	0.20	
F	50	10	0.10	B
D	111	10	0.09	
I	12	1	0.08	
C	223	17	0.08	
B	970	53	0.05	C
E	85	4	0.05	
G	23	1	0.04	
H	17	0	0.00	D

Derived Fields from Text

- Text mining can provide an immense amount of data when limited data may seem to exist.
- Mining the text data may reveal patterns that can be adapted into input variables.
- Text mining is not covered in this course.

Coding in Action

Data Preparation – Categorical Feature Engineering

Data Preparation

Conclusion

- Data Preparation
  - Feature Engineering
  - Recency and Frequency
  - Categorical Feature Engineering

Probability and Statistical Approaches

Probability and Statistical Approaches

- Probability and Statistical Approaches
  - Benford’s Law
  - Z-scores and Robust Z-scores
  - IQR Rule and Its Adjustment
  - Mahalanobis Distances and Robust Mahalanobis

Probability and Statistical Approaches

Benford’s Law

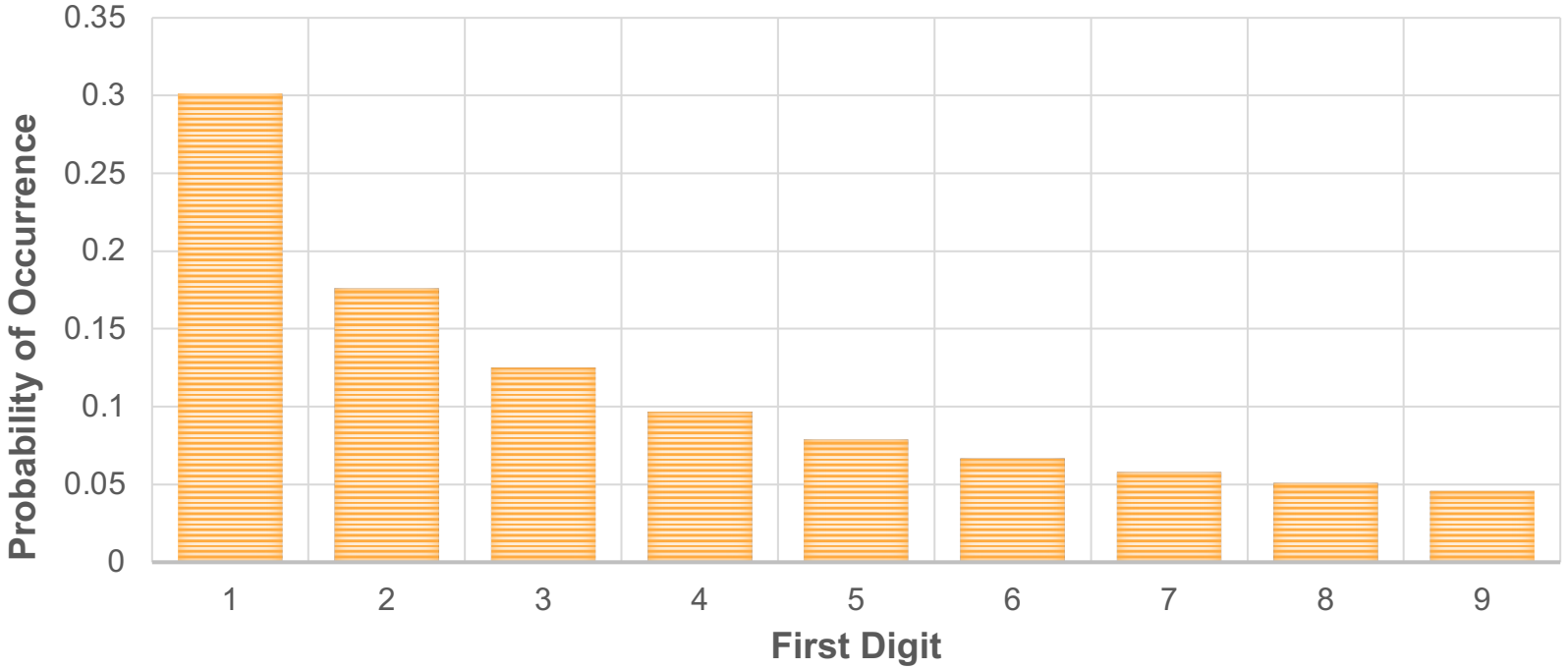
- Probability and Statistical Approaches
  - Benford’s Law
  - Z-scores and Robust Z-scores
  - IQR Rule and Its Adjustment
  - Mahalanobis Distances and Robust Mahalanobis

Benford’s Law

- Certain numbers do not occur uniformly despite what we might think.
- Digits of certain numbers follow Benford’s Law.

- Example:
  - First digit of house/building numbers in addresses.
  - First digit of transaction amounts.

Benford's Law



Benford's Law

- This wasn't mathematically proven until the mid-90's.
- http://testingbenfordslaw.com/

- Benford's Law – First Digit

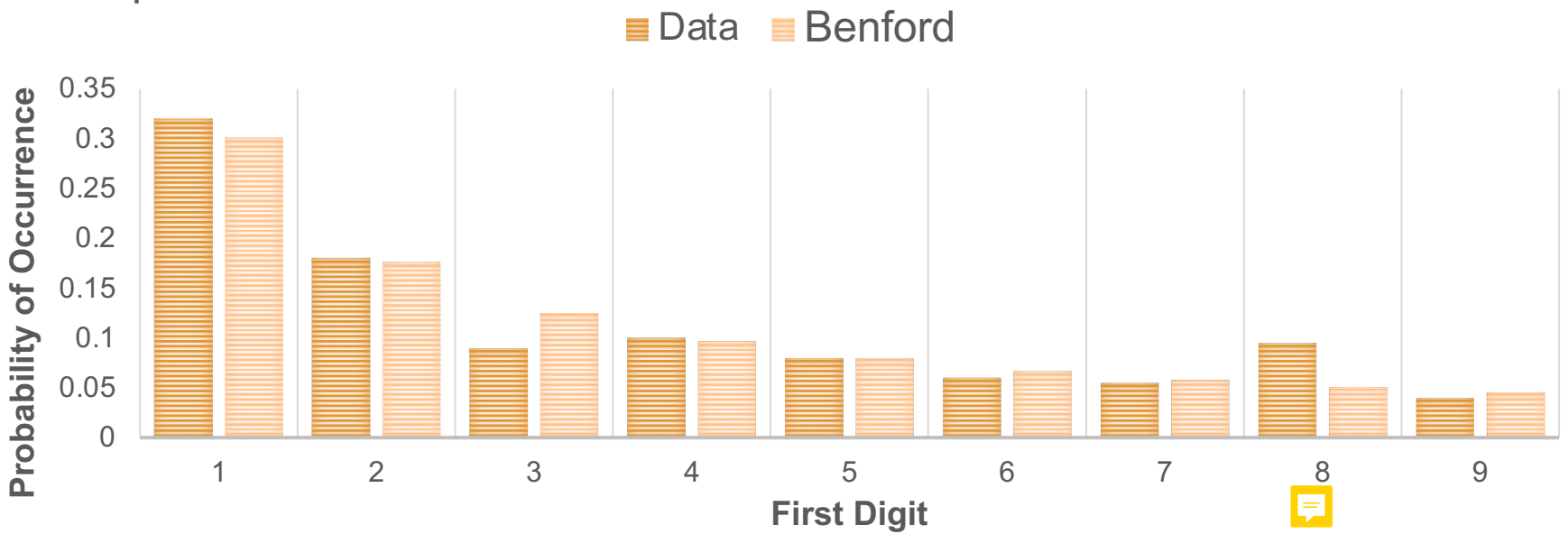
$$P(d_1) = \log_{10}\left(1 + \frac{1}{d_1}\right)$$

Benford’s Law – Fraud Detection

- Fraud transactions typically involve inventing new numbers or changing real transactions into fraudulent ones.
- Legally admissible in Federal, State, and Local courts in United States as evidence.

Benford's Law – Fraud Detection

- Example transaction amounts submitted for reimbursement from scanned receipts



Benford's Law

- Fraud detection typically uses the first two digits in Benford's Law.
- Benford's Law – First Two Digits

$$P(d_1 d_2) = \log_{10}\left(1 + \frac{1}{d_1 d_2}\right)$$

$$d_1 d_2 \in [10,11,12,13, \ldots, 99]$$

Coding in Action

Probability and Statistical Techniques – Benford’s Law

Probability and Statistical Approaches

Z-scores and Robust Z-scores

- Probability and Statistical Approaches
  - Benford's Law
  - Z-scores and Robust Z-scores
  - IQR Rule and Its Adjustment
  - Mahalanobis Distances and Robust Mahalanobis

Statistical Methods

- Basic fraudulent systems look for abnormal observations from a statistical standpoint.
- Univariate analysis can help identify fraudulent **transactions** or **people** (aggregated transactions).

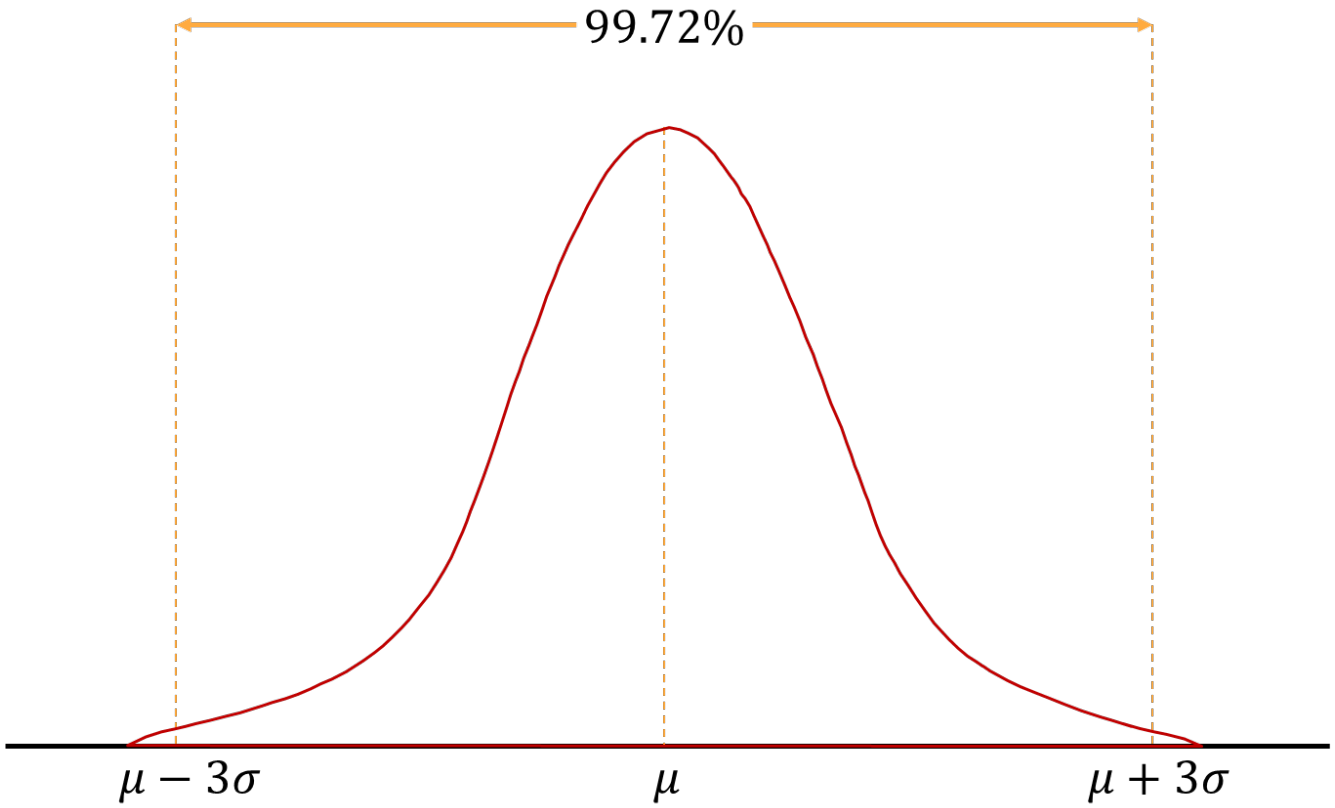
Z-Scores

- Typical with Normal distributions.

$$z_i = \frac{x_i - \bar{x}}{s}$$

- Measures how many standard deviations away from mean each point is.
- Works best with **symmetric** distributions.

Empirical Rule



Z-Scores

- Typical with Normal distributions.

$$z_i = \frac{x_i - \bar{x}}{s}$$

Bothered by outliers

- Measures how many standard deviations away from mean each point is.
- Works best with **symmetric** distributions.

Robust Statistics

- Outliers can greatly influence results.
- Robust techniques

1. Reliable when outliers present
2. Reliable when outliers **not** present (ideally)

Robust Z-Scores

- Robust adjustments to mean and standard deviation.

$$z_{R,i} = \frac{x_i - \text{median(x)}}{\text{MAD}(x)}$$

- Median Absolute Deviation (MAD):

$$\text{MAD}(x) = k \times \text{median}(|x_i - \text{median}(x)|)$$

Robust Z-Scores

- Robust adjustments to mean and standard deviation.

$$z_{R,i} = \frac{x_i - \text{median(x)}}{\text{MAD}(x)}$$

- Median Absolute Deviation (MAD):

$$\text{MAD}(x) = k \times \text{median}(|x_i - \text{median}(x)|)$$

Adjustment factor per distribution

Robust Z-Scores

- Robust adjustments to mean and standard deviation.

$$z_{R,i} = \frac{x_i - \text{median(x)}}{\text{MAD}(x)}$$

- Median Absolute Deviation (MAD):

$$\text{MAD}(x) = k \times \text{median}(|x_i - \text{median}(x)|)$$

1.4826 for Normal distribution

Coding in Action

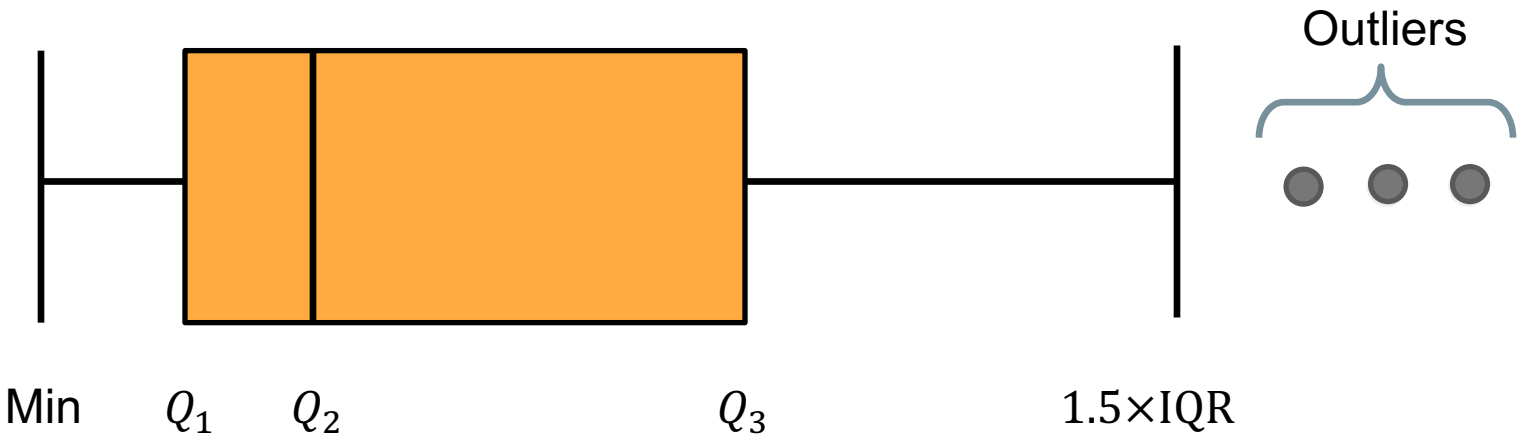
Probability and Statistical Techniques – Z-Scores and Robust Z-Scores

Probability and Statistical Approaches

IQR Rule and Its Adjustment

- Probability and Statistical Approaches
  - Benford’s Law
  - Z-scores and Robust Z-scores
  - IQR Rule and Its Adjustment
  - Mahalanobis Distances and Robust Mahalanobis

1.5 IQR Rule



1.5 IQR Rule

- Works best for **symmetric** distributions.
- Severely skewed distributions tend to report large number of outliers.
- Some software (like R) can use **adjusted boxplot** instead – more robust to skewed distributions.

Coding in Action

Probability and Statistical Techniques – IQR Rule and Its Adjustment

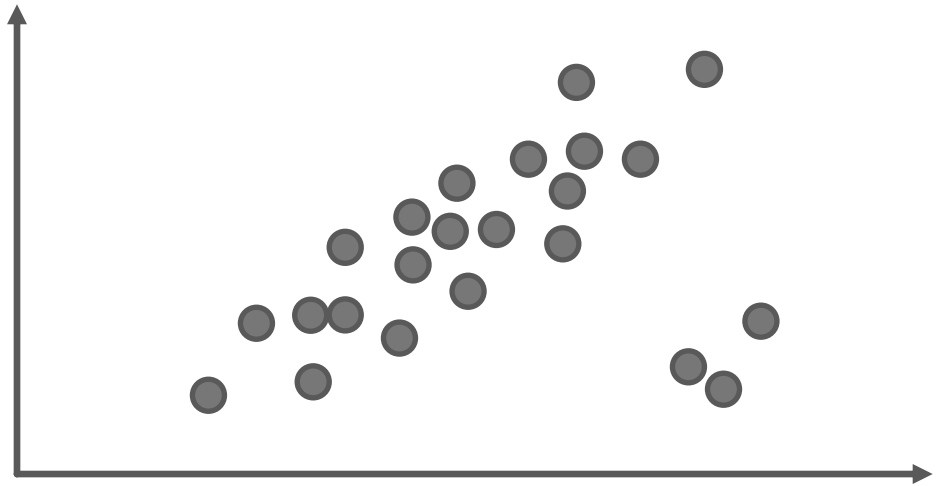
Probability and Statistical Approaches

Mahalanobis Distances and Robust Mahalanobis

- Probability and Statistical Approaches
  - Benford’s Law
  - Z-scores and Robust Z-scores
  - IQR Rule and Its Adjustment
  - Mahalanobis Distances and Robust Mahalanobis

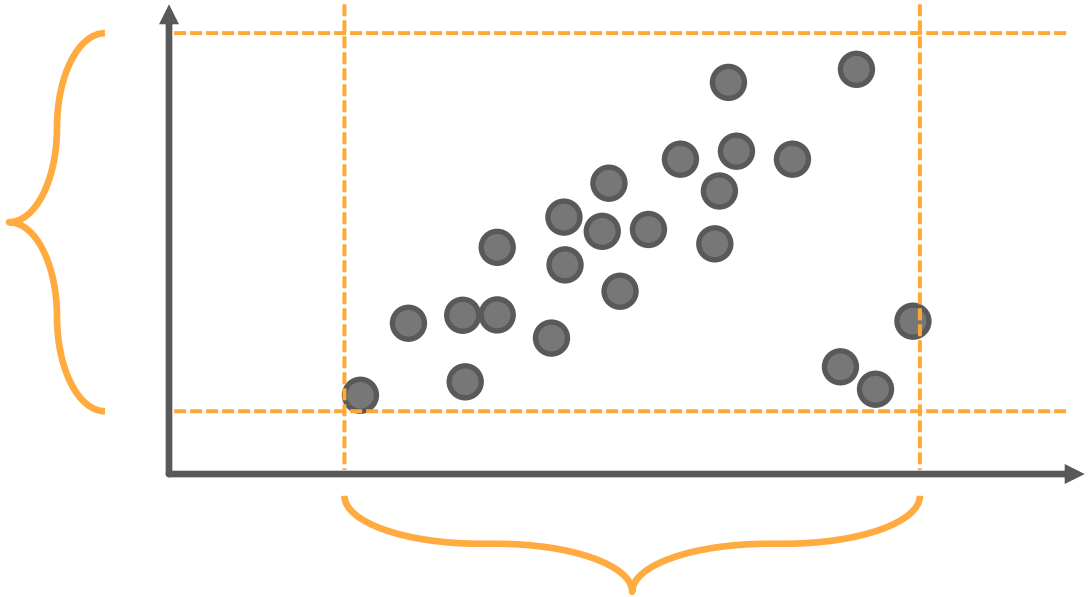
Multiple Dimensions

- Outliers in one dimension are possibly restrictive.



Multiple Dimensions

- Outliers in one dimension are possibly restrictive.



Mahalanobis Distances

- Generalization of z-scores to multi-dimensional space.
  - Replace univariate mean with **multivariate mean**
  - Replace standard deviation with **covariance matrix**

Mahalanobis Distances

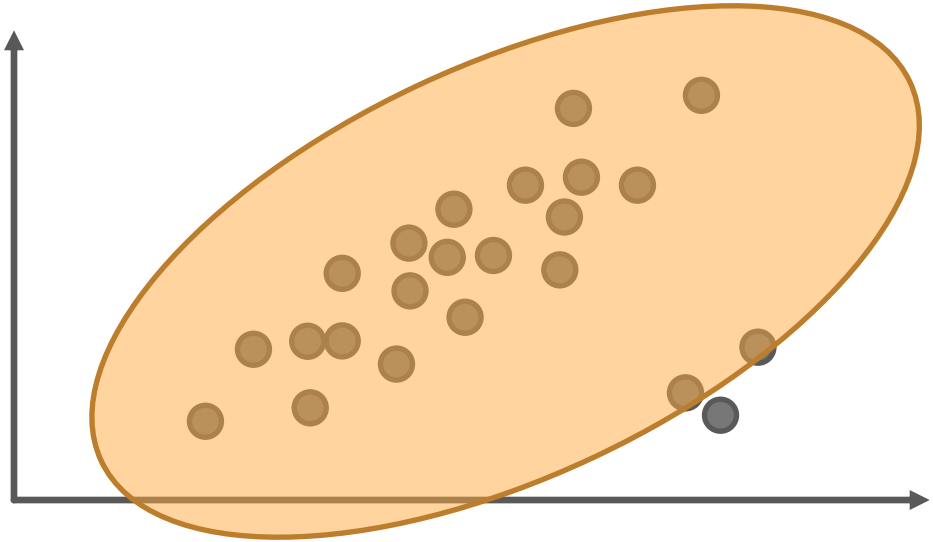
- Generalization of z-scores to multi-dimensional space.
  - Replace univariate mean with **multivariate mean**
  - Replace standard deviation with **covariance matrix**

- Euclidean Distance (L2): $D_{L2} = \sqrt{(x-\mu)^T(x-\mu)}$

- Mahalanobis Distance: $D_M = \sqrt{(x-\mu)^T\Sigma^{-1}(x-\mu)}$

Confidence Ellipsoids

- Still bothered by outliers since standard mean and covariance matrix used.



Robust Mahalanobis Distances

- Mahalanobis distances use mean and covariance matrix influenced by outliers.
- Use **robust** calculations of mean vector and covariance matrix instead:

$$D_M = \sqrt{(x - \mu_{MCD})^T \Sigma_{MCD}^{-1} (x - \mu_{MCD})}$$

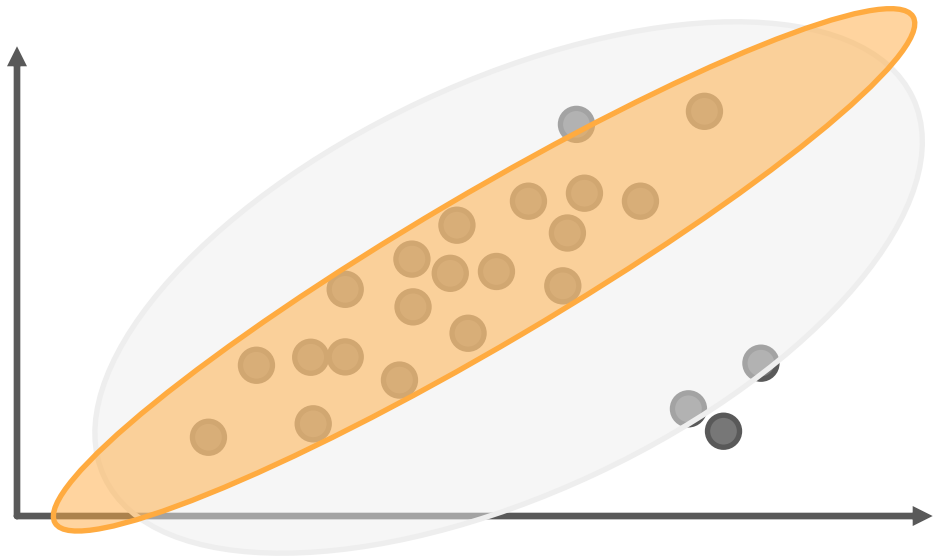
Robust Mahalanobis Distances

$$D_M = \sqrt{(x - \mu_{MCD})^T \Sigma_{MCD}^{-1} (x - \mu_{MCD})}$$

- MCD: Minimum Covariance Determinant
  - Find h $(< n)$ observations that have MCD (essentially the tightest cloud)
  - Typically $h = 0.75 \times n$
  - Problem: How to find the right *h* observations?
  - Fast algorithms exist

Confidence Ellipsoids

- Robust version isn't impacted by outliers as drastically.



Coding in Action

Probability and Statistical Techniques –

Mahalanobis Distance & Robust Mahalanobis

Probability and Statistical Approaches

Conclusion

- Probability and Statistical Approaches
  - Benford’s Law
  - Z-scores and Robust Z-scores
  - IQR Rule and Its Adjustment
  - Mahalanobis Distances and Robust Mahalanobis

Machine Learning Approaches

Machine Learning Approaches

- Machine Learning Approaches
  - k-Nearest Neighbors (kNN)
  - Local Outlier Factor (LOF)
  - Isolation Forests
  - Classifier-Adjusted Density Estimation (CADE)
  - One-Class Support Vector Machine (SVM)

Machine Learning Approaches

k-Nearest Neighbors (kNN)

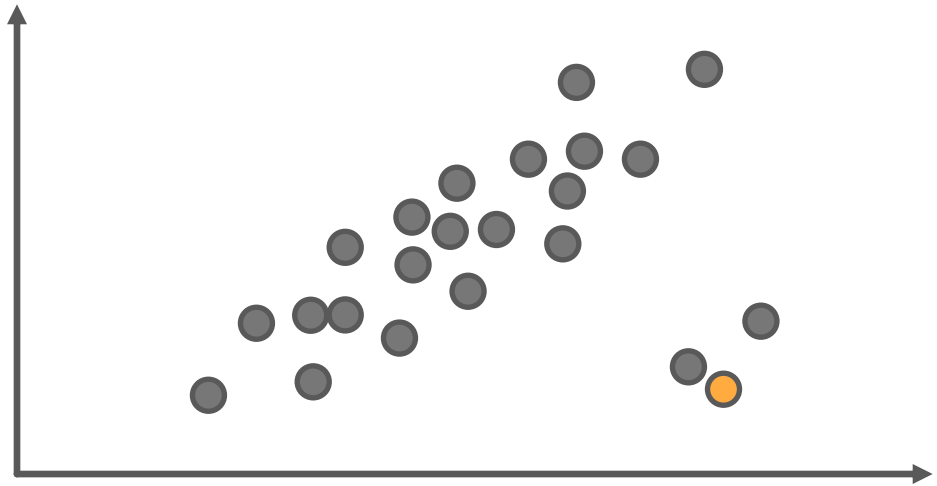
- Machine Learning Approaches
  - k-Nearest Neighbors (kNN)
  - Local Outlier Factor (LOF)
  - Isolation Forests
  - Classifier-Adjusted Density Estimation (CADE)
  - One-Class Support Vector Machine (SVM)

k-Nearest Neighbors

- Want to discover points that are “not close” to the rest.
- Instead of distance from center of cloud, k-NN looks at distance from close points.
- Measure **average** distance from a point to each of the k-closest points.
    - Default: Euclidean distance

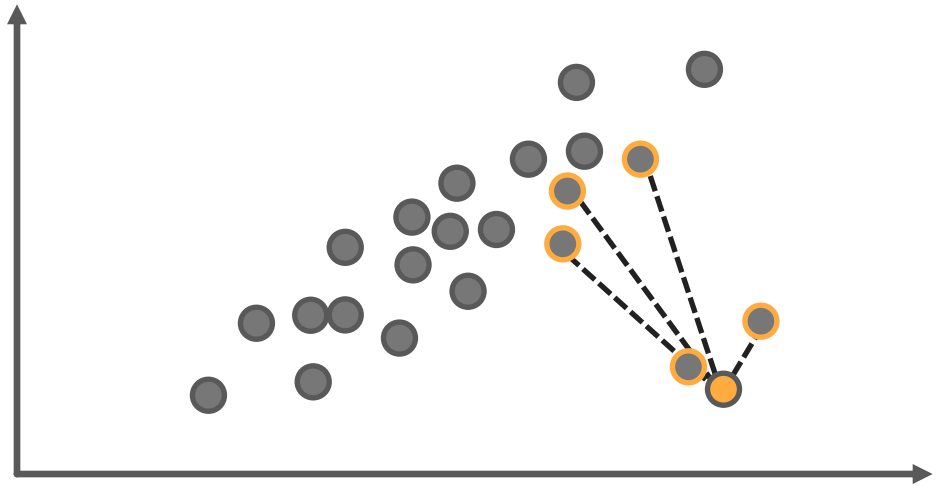
k-Nearest Neighbors

- Need to measure distances to *k* nearest observations.



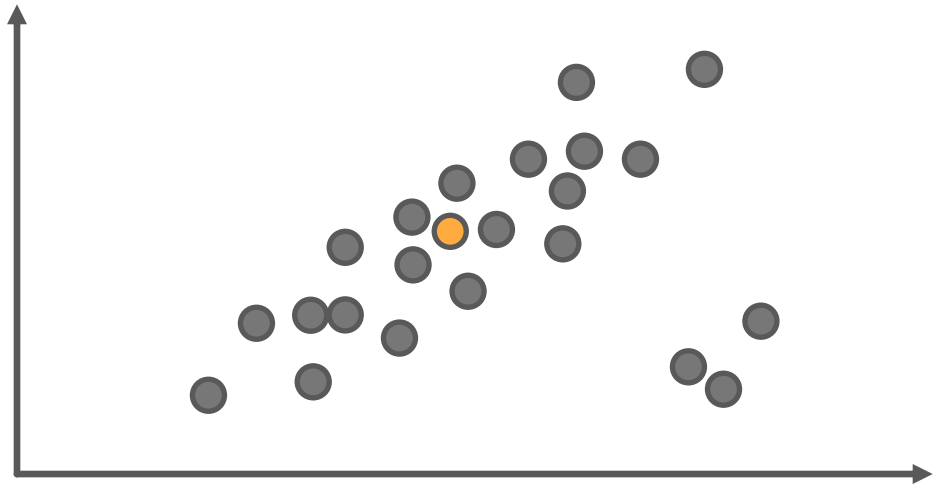
k-Nearest Neighbors

- Need to measure distances to *5* nearest observations.



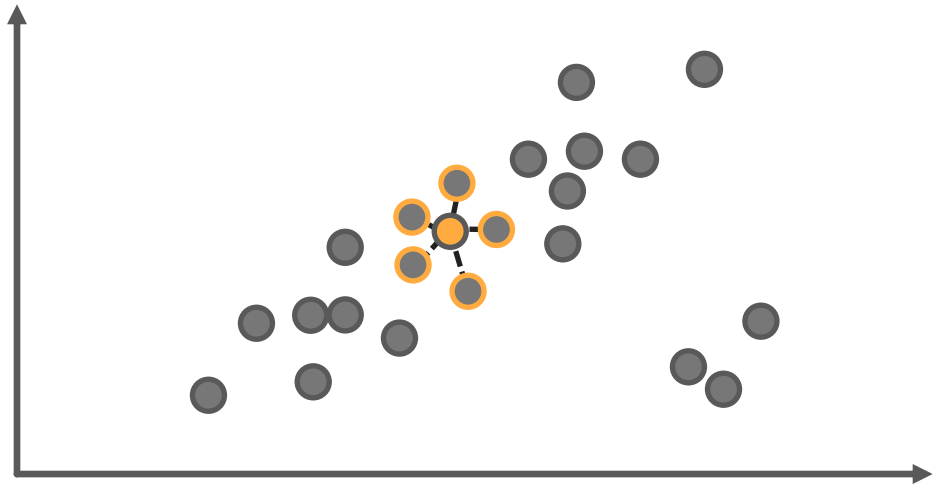
k-Nearest Neighbors

- Need to measure distances to *k* nearest observations.



k-Nearest Neighbors

- Need to measure distances to *5* nearest observations.



Coding in Action

Machine Learning Techniques – k-Nearest Neighbors

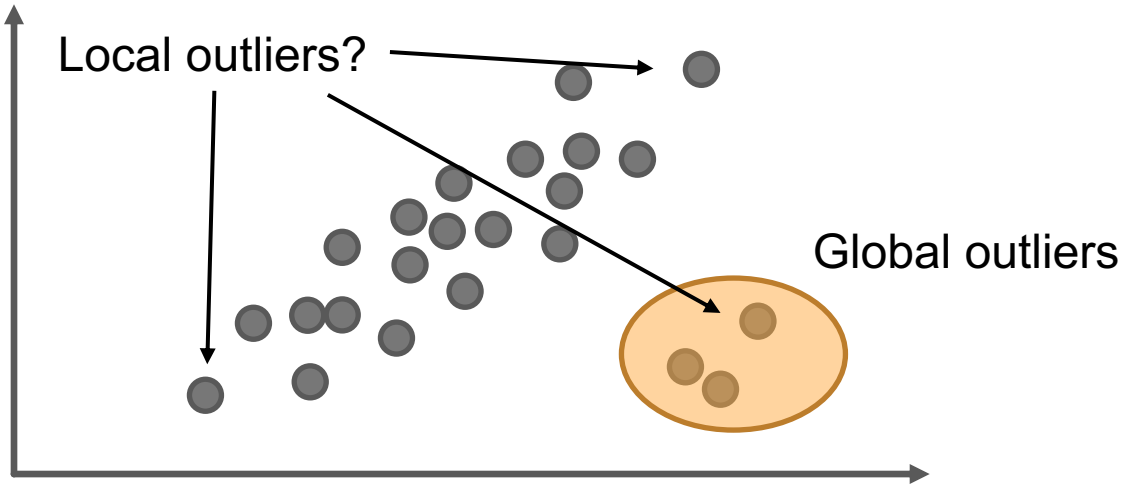
Machine Learning Approaches

Local Outlier Factor (LOF)

- Machine Learning Approaches
  - k-Nearest Neighbors (kNN)
  - Local Outlier Factor (LOF)
  - Isolation Forests
  - Classifier-Adjusted Density Estimation (CADE)
  - One-Class Support Vector Machine (SVM)

Global vs. Local Outliers

- k-NN great at detecting **global** outliers, but not **local** outliers.



Local Outlier Factor (LOF)

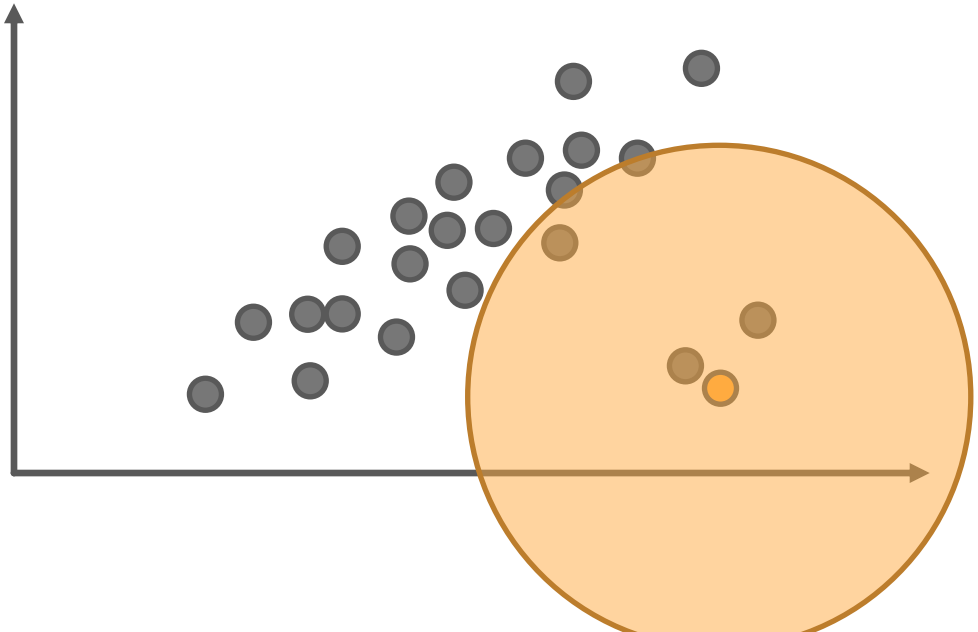
- LOF:
  - Ratio (comparison) of the average **density** of the k-NN of an observation to the **density** of the observation itself.
  - > 1 means more likely to be anomaly
  - < 1 means less likely to be anomaly

Local Outlier Factor (LOF)

- LOF:
    - Ratio (comparison) of the average **density** of the k-NN of an observation to the **density** of the observation itself.
- Density:
    - Inverse of the average **reachability** (distances) from observation to all of its k-NN.
    - Essentially, how far do we have to travel to nearest point, so less dense means farther travel.

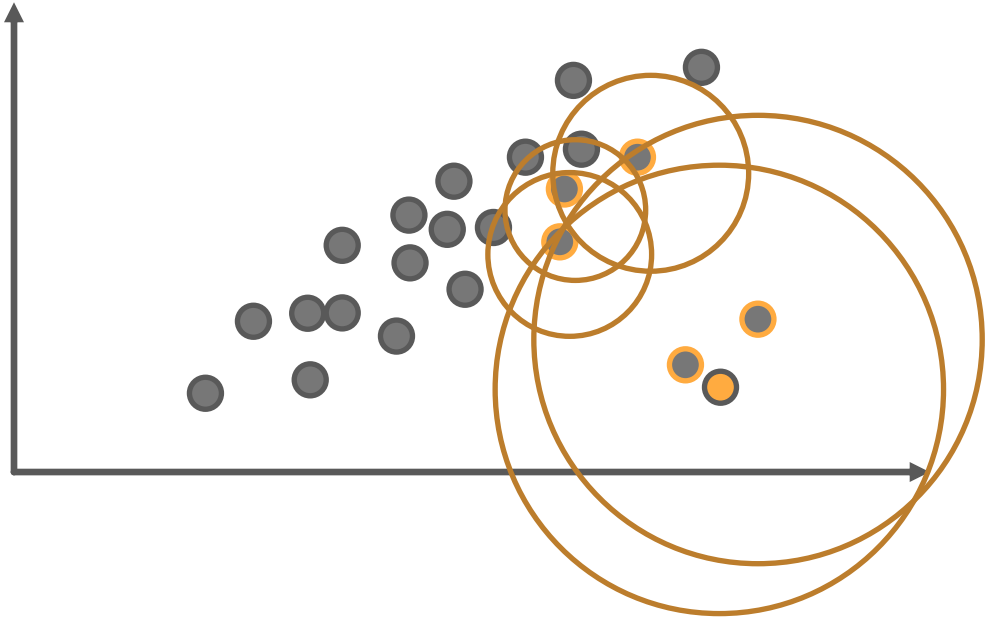
Local Outlier Factor (LOF)

- Density of observation of interest



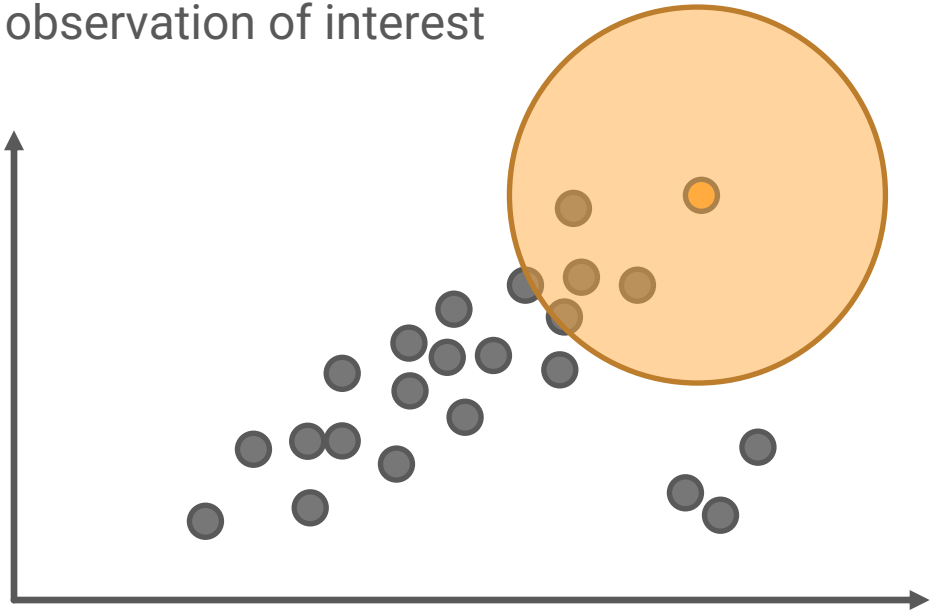
Local Outlier Factor (LOF)

- Need to average the densities of the k-NN observations.



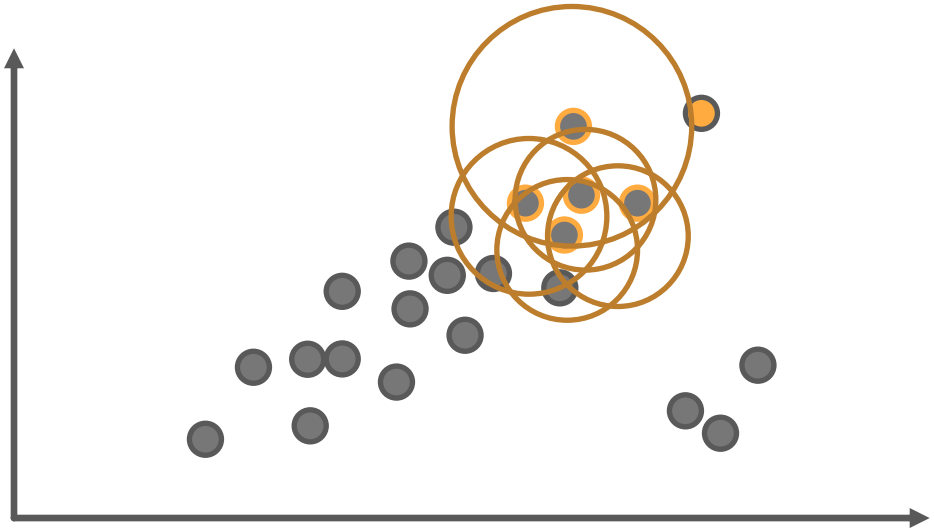
Local Outlier Factor (LOF)

- Density of observation of interest



Local Outlier Factor (LOF)

- Need to average the densities of the k-NN observations.



Coding in Action

Machine Learning Techniques – Local Outlier Factor (LOF)

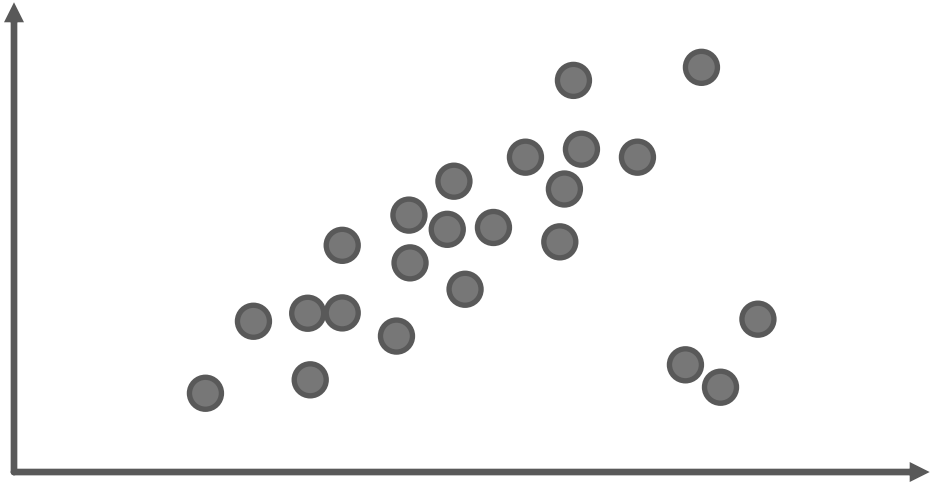
Machine Learning Approaches

Isolation Forests

- Machine Learning Approaches
  - k-Nearest Neighbors (kNN)
  - Local Outlier Factor (LOF)
  - Isolation Forests
  - Classifier-Adjusted Density Estimation (CADE)
  - One-Class Support Vector Machine (SVM)

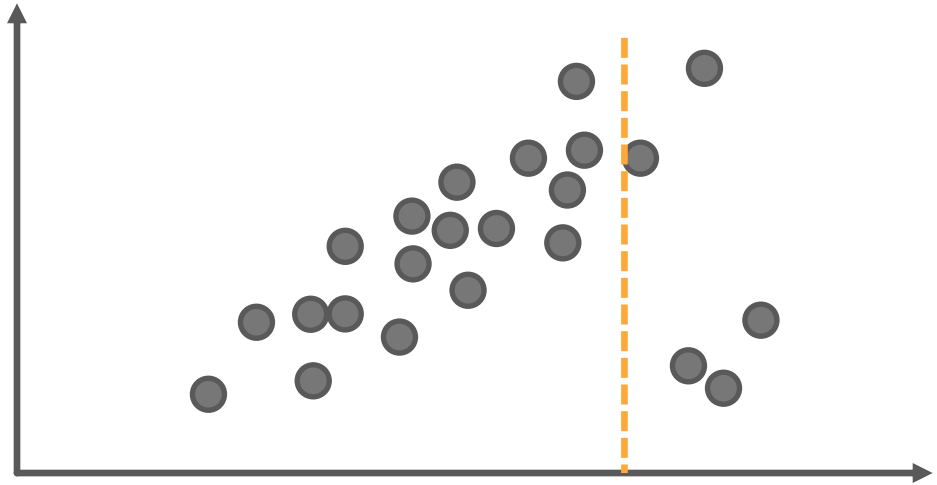
Isolation Tree

- Tree-based algorithm to isolate observations.
- Easier the isolation → More likely an anomaly!



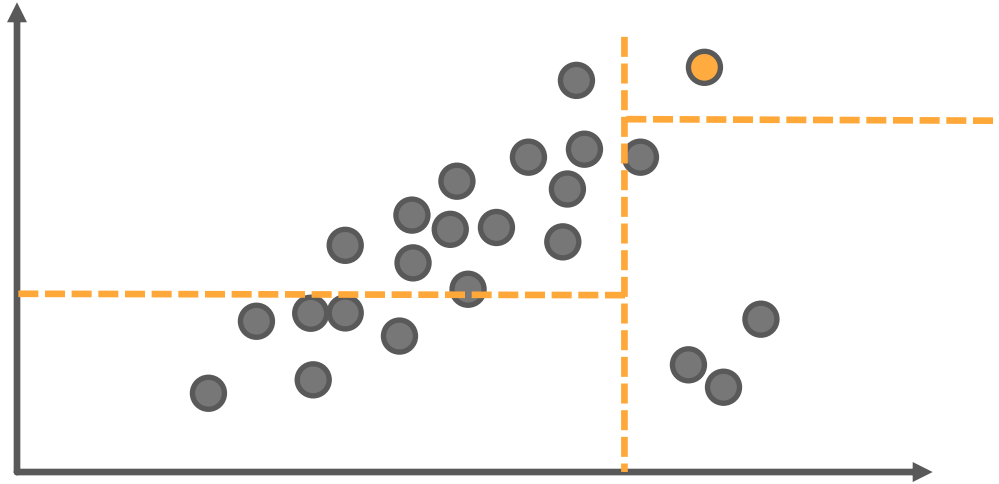
Isolation Tree

- Tree-based algorithm to isolate observations.
- Easier the isolation → More likely an anomaly!



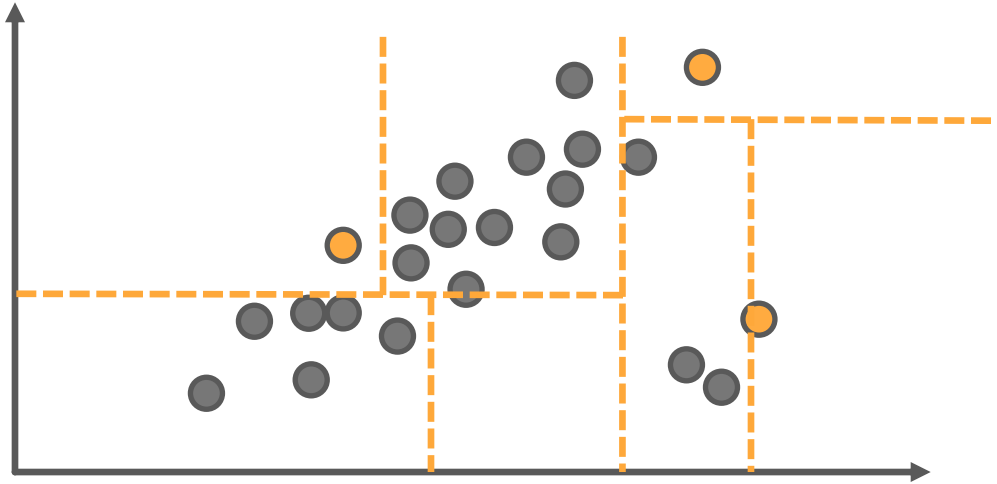
Isolation Tree

- Tree-based algorithm to isolate observations.
- Easier the isolation → More likely an anomaly!



Isolation Tree

- Tree-based algorithm to isolate observations.
- Easier the isolation → More likely an anomaly!



Isolation Tree

- Tree-based algorithm to isolate observations.
- Easier the isolation → More likely an anomaly!
- Isolation score is inversely related to number of needed splits to isolate observation.
    - Bounded between 0 and 1.
    - Closer to 1 → more likely an anomaly
    - Closer to 0 → less likely an anomaly
    - All observations ~ 0.5, no real anomalies

Isolation Forest

- Since the isolation trees are based on random splits on random dimensions, outlier might get lucky and survive longer than it really should.
- Isolation forest – combination of MANY isolation trees with averaged scores.
- Look for convergence of scores for optimal number of trees.

Coding in Action

Machine Learning Techniques – Isolation Forest

Machine Learning Approaches

Classifier-Adjusted Density Estimation (CADE)

- Machine Learning Approaches
  - k-Nearest Neighbors (kNN)
  - Local Outlier Factor (LOF)
  - Isolation Forests
  - Classifier-Adjusted Density Estimation (CADE)
  - One-Class Support Vector Machine (SVM)

CADE

- Newer technique for density estimation.
- Value been found in anomaly detection and fraud applications.

CADE

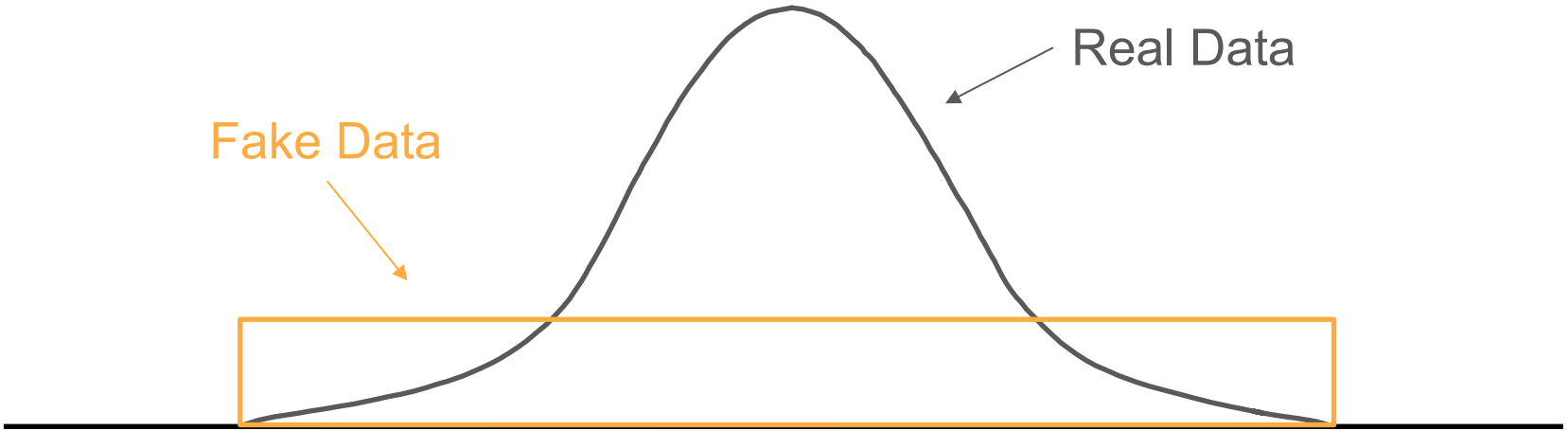
- Process:
  1. Label all original data as **not outliers**
  2. Create new observations (same *n* as data) but variables are all uniformly distributed
  3. Label all new data as **outliers**, merge old and new data
  4. Use classification model to predict “outliers” (1’s).
  5. Score original data

CADE

- High predicted probabilities → More likely an anomaly!
- Observation looks more like fake uniform data than actual distribution from which it came in multivariate space.

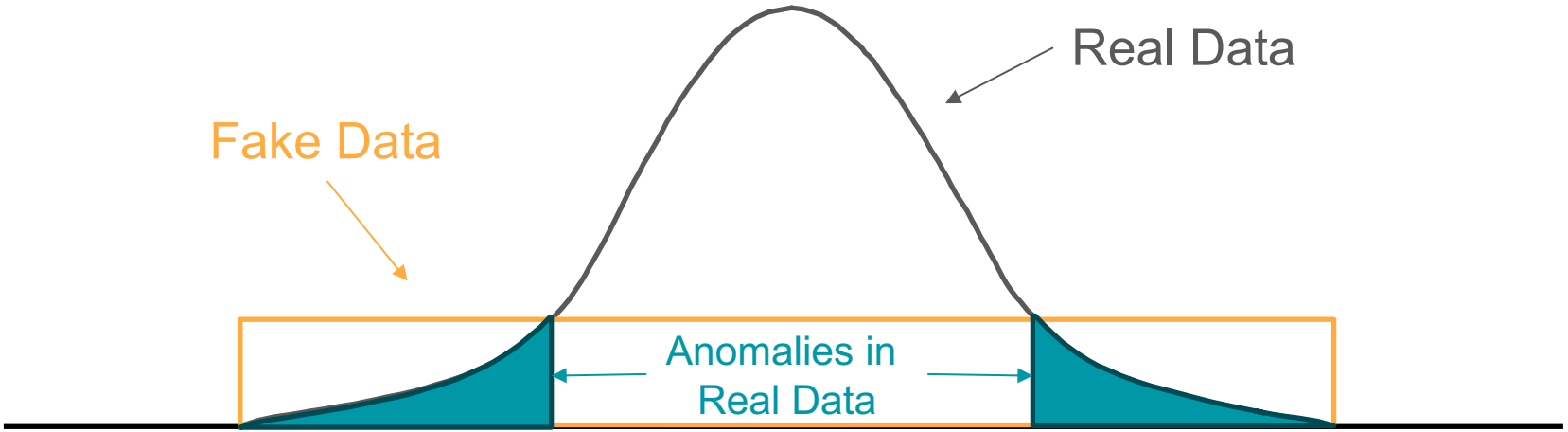
CADE

- High predicted probabilities → More likely an anomaly!
- Observation looks more like fake uniform data than actual distribution from which it came in multivariate space.



CADE

- High predicted probabilities → More likely an anomaly!
- Observation looks more like fake uniform data than actual distribution from which it came in multivariate space.



Coding in Action

Machine Learning Techniques – Classifier-Adjusted Density Estimation

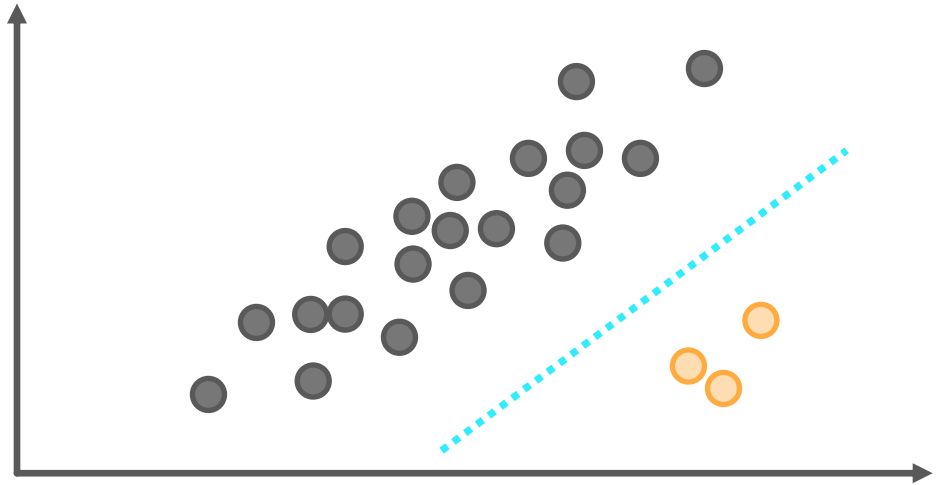
Machine Learning Approaches

One-Class Support Vector Machine (SVM)

- Machine Learning Approaches
  - k-Nearest Neighbors (kNN)
  - Local Outlier Factor (LOF)
  - Isolation Forests
  - Classifier-Adjusted Density Estimation (CADE)
  - One-Class Support Vector Machine (SVM)

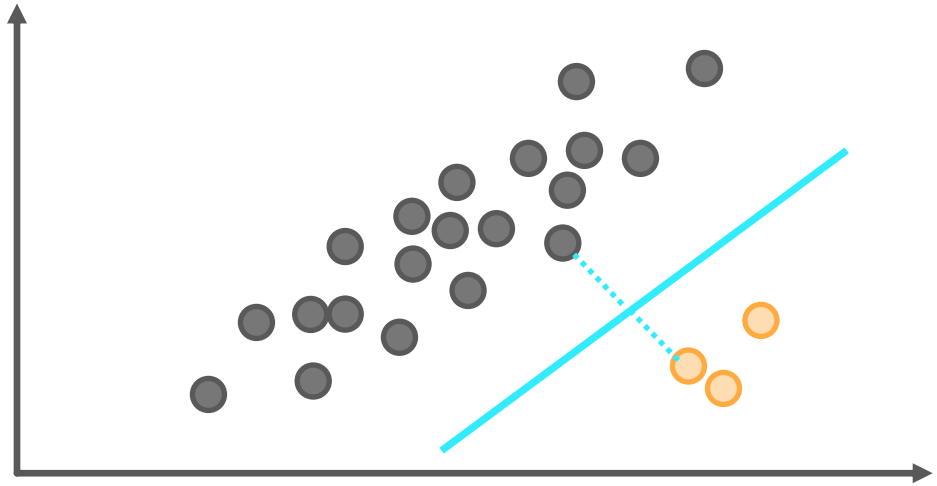
Support Vector Machines (SVM)

- A traditional two-class SVM is a classifier.
- It creates a hyperplane that “best” separates the two classes.



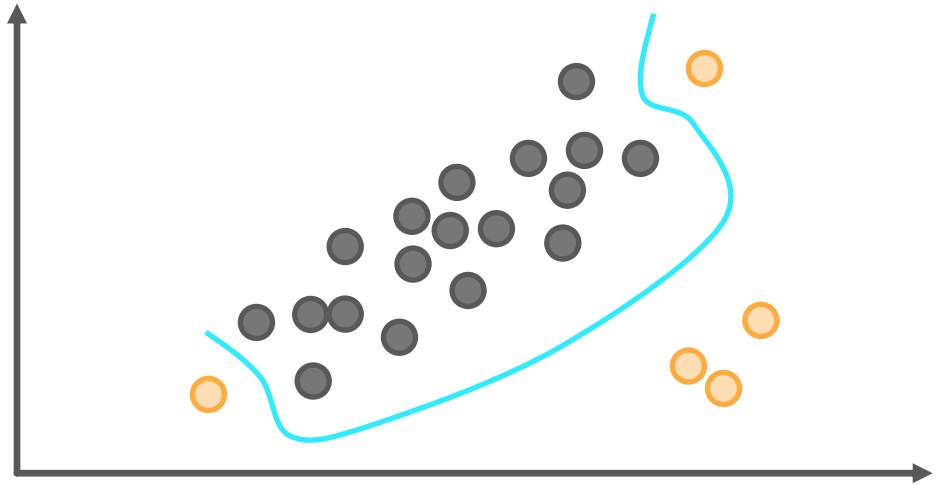
Support Vector Machines (SVM)

- A traditional two-class SVM is a classifier.
- It creates a hyperplane that “best” separates the two classes.
- “Best” is maximizing the distance (in every dimension) from the two classes.



Support Vector Machines (SVM)

- A traditional two-class SVM is a classifier.
- Not limited to linear separation! Kernels are used to make hyperplanes nonlinear.



One-Class Support Vector Machines

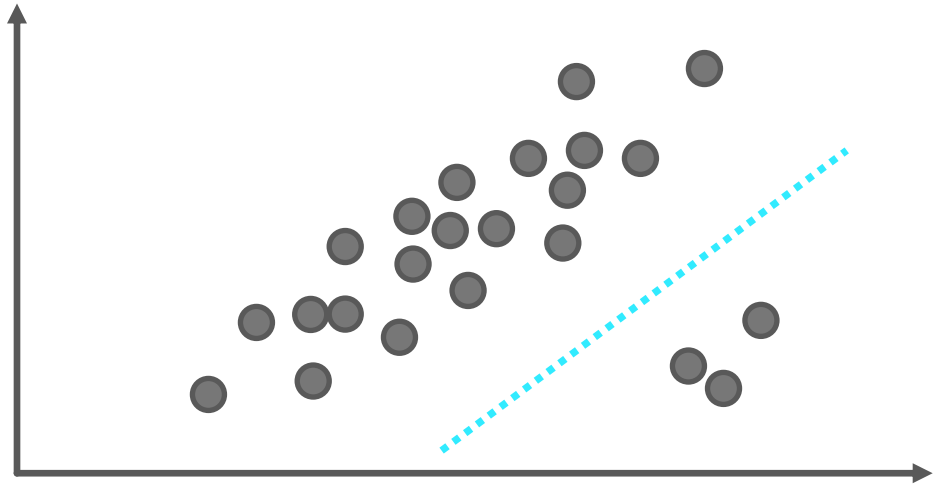
- SVM's are also used in an unsupervised learning scenario as well.
- Instead of thinking about two classes, we can take one of two approaches:

1. Tell the SVM to isolate the X% of observations. If we think we have 5% anomalies, we tell the SVM to isolate the "most" anomalous 5% of observations.

2. Train the SVM on all data as normal and score new data to see if it falls "within normal".

One-Class Support Vector Machines

- Tell the SVM to isolate the X% of observations. If we think we have 5% anomalies, we tell the SVM to isolate the “most” anomalous 5% of observations.



Coding in Action

Machine Learning Techniques – One-Class Support Vector Machine

Machine Learning Approaches

Conclusion

- Machine Learning Approaches
  - k-Nearest Neighbors (kNN)
  - Local Outlier Factor (LOF)
  - Isolation Forests
  - Classifier-Adjusted Density Estimation (CADE)
  - One-Class Support Vector Machine (SVM)

Conclusion

Course Outline

- Introduction
- Data Preparation
- Probability & Statistical Techniques
- Machine Learning Techniques
- Conclusion

Course Outline

- Introduction
    - Who am I?
    - What are Anomalies?
    - Anomaly Detection Analytical Framework
- Data Preparation
- Probability & Statistical Techniques
- Machine Learning Techniques
- Conclusion

Course Outline

- Introduction
- Data Preparation
    - Feature Engineering
    - Recency and Frequency
    - Categorical Feature Engineering
- Probability & Statistical Techniques
- Machine Learning Techniques
- Conclusion

Course Outline

- Introduction
- Data Preparation
- Probability & Statistical Techniques
    - Benford’s Law
    - Z-scores and Robust Z-scores
    - IQR Rule and Its Adjustment
    - Mahalanobis Distances and Robust Mahalanobis
- Machine Learning Techniques
- Conclusion

Course Outline

- Introduction
- Data Preparation
- Probability & Statistical Techniques
- Machine Learning Techniques
    - k-Nearest Neighbors (k-NN)
    - Local Outlier Factor (LOF)
    - Isolation Forests
    - Classifier-Adjusted Density Estimation (CADE)
    - One-Class Support Vector Machine (SVM)
- Conclusion

Where Am I?

- Find me online:
  - https://www.linkedin.com/in/ariclabarr/
  - https://www.youtube.com/c/AricLaBarr/
  - https://www.ariclabarr.com/

Thank you

TRAINING