



DEEP
LEARNING
INSTITUTE



DLI Accelerated Data Science Teaching Kit

Lecture 6.2 - Break Complex Problems into Simpler Ones: Part 2



The Accelerated Data Science Teaching Kit is licensed by NVIDIA, Georgia Institute of Technology, and Prairie View A&M University under the [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

5. Co-occurrence grouping

(Many names: frequent itemset mining, association rule discovery, market-basket analysis)

Find associations between entities based on transactions that involve them

(e.g., bread and milk often bought together)



How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

<http://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/>

6. Profiling/Pattern Mining/Anomaly Detection

Characterize typical behaviors of an entity (person, computer router, etc.)
so you can find trends and outliers.

- Google sign-in alert
- Computer instruction prediction
- Removing noisy data (data cleaning)
- Detect anomalies in network traffic
- Moneyball
- Smart security camera

7. Link Prediction/ Recommendation

Predict if two entities should be connected, and how strongly that link should be.

Linkedin/Facebook: people you may know

Amazon/Netflix/Pandora: because you like terminator...suggest other movies you may also like

8. Data Reduction

(“dimensionality reduction”)

Shrink a large dataset into smaller one, with as little loss of information as possible

1. if you want to visualize the data (in 2D/3D)
2. faster computation/less storage
3. reduce noise



DEEP
LEARNING
INSTITUTE



DLI Accelerated Data Science Teaching Kit

Thank You