IS4242
INTELLIGENT SYSTEMS & TECHNIQUES

L4 – Finding New Customers

Aditya Karanam

Announcements

- Programming Assignment – 1, Due: September 12, 11:59 PM
  - Any issues with the submission?

- Quizzes: Test your conceptual understanding of business problems and ML techniques
  - You will not be asked to derive

- Correction in the formula for Precision in the previous class
  - Please use the updated slide deck

In this Class …

- Consumer Acquisition
- Unsupervised Learning
  - K-Means
  - Agglomerative Clustering
- Dimensionality Reduction
  - Principal Component Analysis

Finding New Customers

- More than half of US startups were unable to survive beyond three years from 2001-2010

- Harder to scale the company beyond certain levels of sale
  - E.g., Among 40K companies founded and listed in S&P Capital IQ database, <6% achieved >$10M by 2010 and <2% grew >$50M

- Complexity greatly increases once a venture reaches a critical stage
  - Lot of *interdependent and moving* parts, making it difficult to manage
  - SG&A (selling, general and administrative) expenses accelerate faster than revenues

- Promising ventures *cannot afford* to burn through working capital
  - Either go out of business or operate in small niches

https://sloanreview.mit.edu/article/how-to-identify-the-best-customers-for-your-business/

Finding New Customers: Example

- Pets.com: online retail store to buy pet supplies
- Spent an acquisition cost of $400 per customer
  - Heavily invested into advertising: tv, radio, newspaper, etc.
- Profit potential of acquiring a pet food customer is too low!
  - Consider that average customer spends $100 per purchase and the incremental profit margin is $20
  - Assuming no retention costs, 20 purchases before customer breaks even
- During its first fiscal year (February to September 1999), Pets.com earned $619,000 in revenue, and spent $11.8 million on advertising.
  - Eventually, the biggest disaster of the dot-com era!

https://www.mckinsey.com/capabilities/strategy-and-corporate-finance/our-insights/five-ways-cfos-can-make-cost-cuts-stick

Finding New Customers

- Belief among companies:
  - Acquisition costs would drop significantly, and earn profits from retained customers
- Reality:
  - Almost none of them can drive down their acquisition costs to a level that made it feasible to make a profit.
- Between 2000 – 2010, the cost of goods sold at the *average S&P 500 company* reduced by 250% while SG&A as a percentage of revenue didn't change
- Cost of customer acquisition can sink a business, especially in this digital world!

Customer Equity

- Customer '*equity*' to the firm is made up of two factors:
  - Initial benefits of customer acquisition
  - Profits from customer retention

- Benefits of customer acquisition can be formulated as:

$$N_t \alpha_t (S_t - C_t) - N_t B_t$$

  - N_t: Number of customers targeted, B_t: cost of reaching out to a customer, α_t: acquisition rate, $S_t - C_t$: profit margin of the sales

- We can devise several strategies for consumer acquisition

Several Strategies for Customer Acquisition

- Initial Benefits of Customer Acquisition: $N_t \alpha_t (S_t - C_t) - N_t B_t$
  - Increasing market size or potential customers of the product (N_t)
  - Increasing Acquisition Expenditures: B_t
  - Increasing Acquisition Pricing and Promotions: $S_t - C_t$

Increasing Market Size

- Number of potential customers in the market can be increased in several ways
- Suggest or develop new usage occasions
  - Arm & Hammer Baking Soda being used as a *deodorizer* and not just for *baking*
  - American Express allows customers to withdraw cash while they are traveling
- Target New customer segments
  - Costco started off by focusing on *small businesses*, later opened its warehouse club to *individuals* who were willing to pay an annual fee
  - Whole Foods began as an *organic, natural grocery store* but has expanded to "*foodies*" who enjoy higher quality products and perceive the organic benefit as positive.

Increasing Market Size: Disadvantages

- The consumer becomes *confused* about the *positioning* of the brand
  - As BMW expands the reach of its brand with *lower-end offerings* (1 series in Europe), it risks lack of *exclusivity* and damaging its image.
  - The ultimate example is Cadillac, which created the Cimarron, a "juiced-up" version of the Chevrolet Cavalier in early 1980s
    - It hurt the brand image of Cadillac because consumers started perceiving Cadillac as non-luxury vehicle

Increasing Market Size: Summary

- Significant risks are associated with expanding the market through reaching new segments
- Market expansion can risk reducing the acquisition rate (α)
  - The broader the market is, the lower the acquisition rate
- Lower acquisition rate implies less returns on the invested capital, in turn leads to negative profits

Increasing Acquisition Expenditures

- Increasing acquisition expenditures helps in increasing costumer acquisition in two ways:
  - Investing to generate *awareness*
  - Investing in *lead products* to draw consumers to the company

Increasing Acquisition Expenditures: Awareness

- By investing in *advertising*: TV, Radio, Google AdWords, Social Media, etc.
  - Clicks lead to website visits and ultimately sales

- By generating positive *word-of-mouth*
  - Positive word-of-mouth generates awareness and impacts the consumer's intention to purchase
    - E.g.: Restaurants, movies and other entertainment services use word-of-mouth as a key source
  - *Influencer Marketing:* Marketing companies pay influencers in social media (Instagram, YouTube, etc.) to generate credible word-of mouth

Increasing Acquisition Expenditures: Lead products

- Use lead products to acquire customers, and sell other products/services after acquisition to obtain benefits

  - Insurance companies attracted customers with accidental death and dismemberment insurance policy

  - Grocery stores use Coke and Pepsi to bring customers into the store with the goal of selling related products and building the size of the basket

Increasing Acquisition Expenditures: Challenges

- *Acquisition* increases with higher expenditures, but the problem is *pay out!*

- Firms acquire customers at a loss, and *generally* make up the loss on future purchases

- Accounting *Distortion*:
  - The cost is recorded in the acquisition period
  - Revenue are obtained from future purchases: recorded in the retention period
    - Making the acquisition period's profits look worse

- Short-term loss from customer acquisition period leads to under-investment

Acquisition Pricing and Promotions

- As the price *decreases*, acquisition rates almost always *increase*
  - Same with promotions, they can increase the acquisition rate

- However, this strategy can impact future purchases
  - Customers develop *expectations* about the firm's pricing and then evaluate future prices
    - More aggressive the introductory promotion, the lower will be the renewal rate unless the firm again offers an aggressive discount.

- Pricing and promotion also influence which segments of customers are acquired
  - Lower prices likely attract price sensitive consumers

Customer Segmentation

- Interwoven in all these strategies is the identification of potentially viable segments to target

- Greater the firm's *ability* in identifying relevant segments, higher the acquisition response rate

- More important, recognize reasonably *finer* segments to target
  - If the segments are broadly defined (e.g., men 18–49), better to mass market than target

Customer Segmentation Methods for Acquisition: Profiling

- Profile existing customer base and use their characteristics to target potentially new customers.
  - Cluster current customers based on various characteristics like demographics
  - Target new customers with similar characteristics as the current customers

- Clusters should answer two basic questions:
  - '*Who*' the potential new customers are and '*where*' to reach them
  - Economic interpretability along with statistical validation is the key

Disadvantages of Profiling

- Profiling is 'self-fulfilling'
  - Firm only targets customers similar to its current customers
  - Current customers were acquired using similar methods and advertising vehicles

- By targeting customers with similar profiles, the firm never learns if it can attract other types of customers

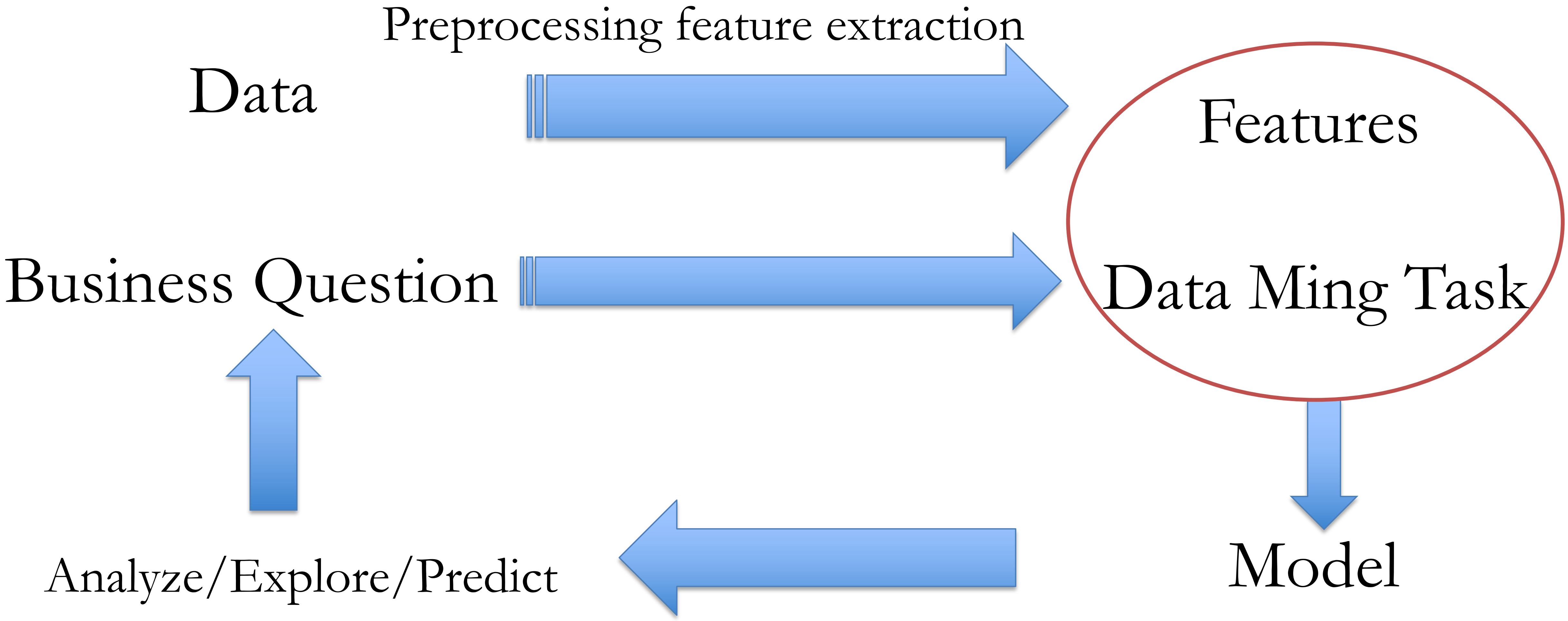
- How to overcome this problem?
  - Use different data for profiling – likely identifies different segments
    - Social media, traveling information along with demographics
  - Different vehicles for targeting – influencer marketing, mobile apps, etc.

Techniques for Segmenting Current Customers for Acquiring New Customers

Application: Marketing Campaign

- Data on 29 attributes of 2240 existing customers
- *Demographics:* Birth year, Education, Marital Status, Income, Children(kids, teens), etc.
- *Products:* Amount spent on different products in the last 2 years
  - Fruits, Meat, Fish, Sweet, Wine, Gold, etc.
- *Place:* Web visits, purchases on web, store, etc.
- *Response:* responses to various campaigns

Data Mining


Preprocessing feature extraction
Data
Features
Business Question
Data Ming Task
Model
Analyze/Explore/Predict

Variable Selection

- We need to know *who* they are and *where* to reach them
  - Demographic and place (media habits) features

- Response variables distort the explanation of clusters
  - We use them to evaluate the segments
  - You can use response variables to identify input variables

- Selecting variables is quite subjective
  - Interpretability or explainability of your results is the key factor

Unsupervised Learning

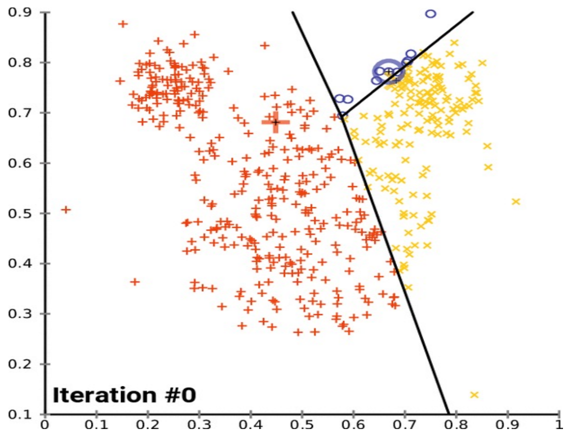
- With the existing consumer data, we cannot verify how the new customer would respond to the targeted marketing campaigns
  - We do not have information on the Y variable (or label)
    - Unsupervised learning task

- Aim: Group or Cluster consumers based on their similarities
  - Customers in the same group are more 'similar' than the customers across groups
    - K-Means Clustering
    - Agglomerative Clustering

K-Means Clustering

- Find K clusters in data (K: input)
- Each cluster associated with a "cluster center" called centroid
- Relies on distance function between feature vectors
  - Distance function is application-dependent
  - Euclidean distance commonly used for continuous-valued vectors
    - Each observation is a vector in p-dimensional space

Lloyd's Algorithm



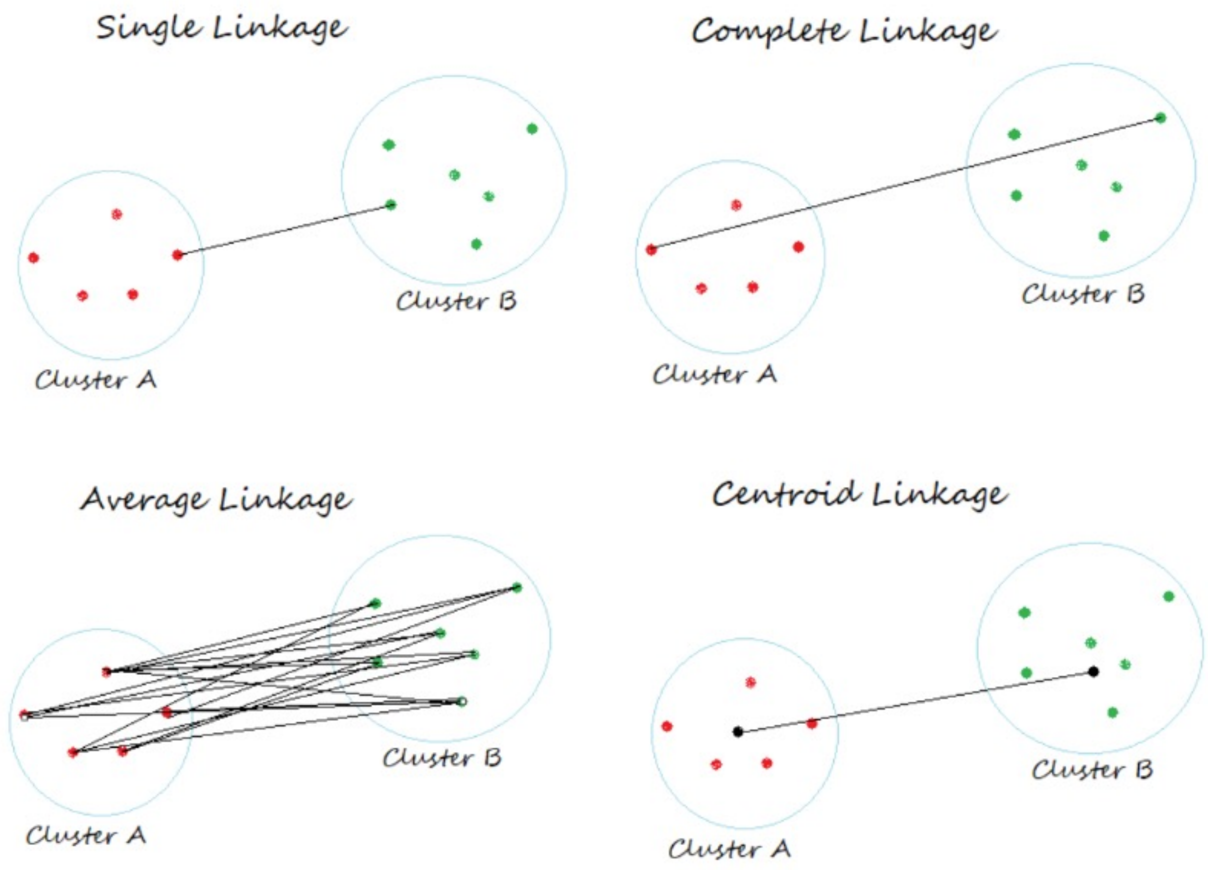
- Input: Data Matrics, K
- Initialization: choose K (random) centroids
  - Note: Results can vary with different choices
- Repeat steps 1 & 2 until convergence:
  1. Find closest centroid to each data point
     - Each data point belongs to the cluster corresponding to the closest centroid
  2. Update centroid
     - In each (current) cluster, the mean of the data points is made the new centroid
- Convergence: None of the data points change their cluster membership
- Output: K clusters of the data
  - Centroids may not be one of the data points

Agglomerative Clustering

- Bottom-up approach
  - Each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy.
- Relies on the distance and the linkage criterion
  - Distance: Euclidean distance for continuous data points
    - Determines which observations are most similar
  - Linkage criterion: influences the distance between clusters or sets of observations

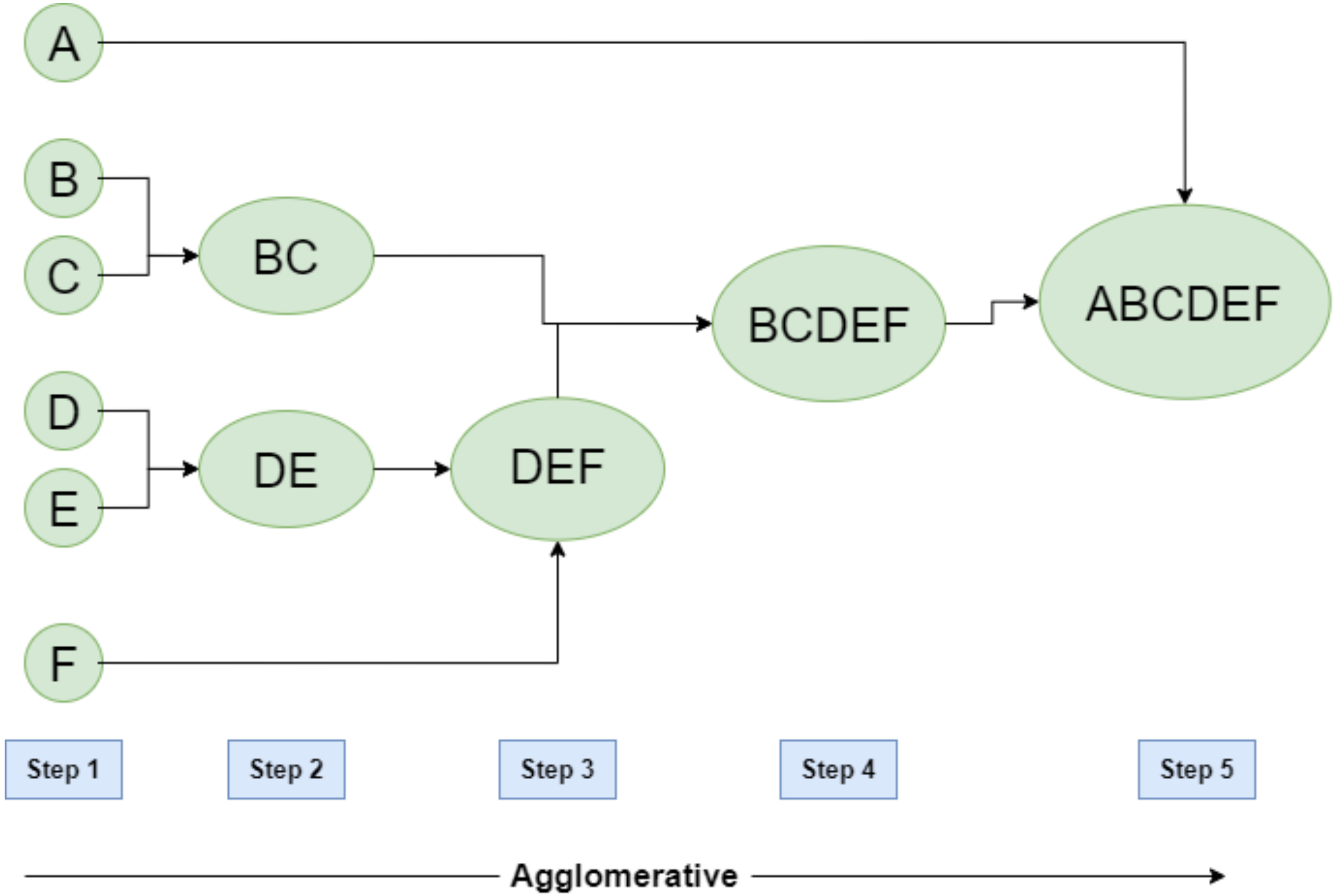
Agglomerative Clustering: Linkage

- Single linkage: Distance between clusters is the minimum distance among members of different clusters
- Complete linkage: Distance between clusters is the maximum distance
- Average Linkage: Distance between clusters is the average distance between members'
- Centroid Linkage: Distance between the cluster centers



https://www.analyticsvidhya.com/blog/2021/06/single-link-hierarchical-clustering-clearly-explained/

Agglomerative Clustering: Example



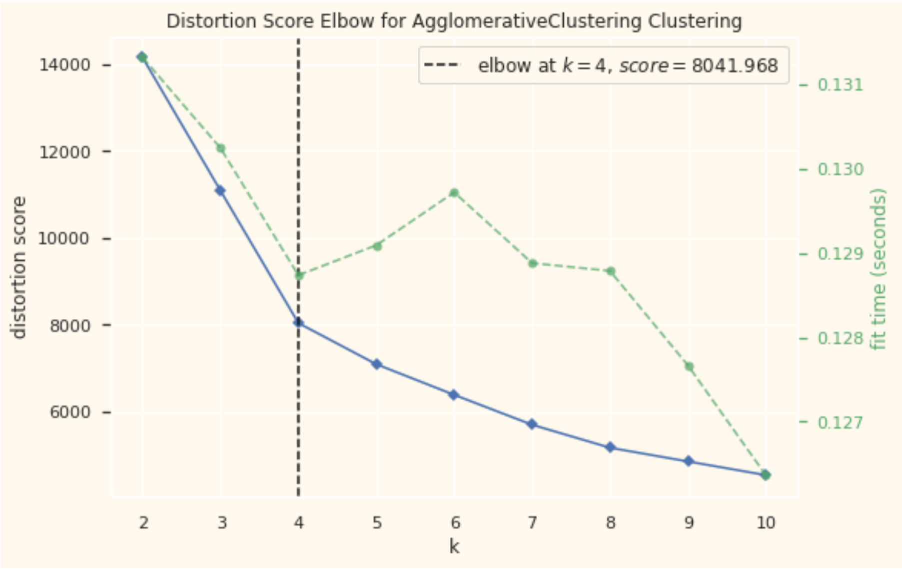
https://www.geeksforgeeks.org/ml-hierarchical-clustering-agglomerative-and-divisive-clustering/

Evaluation

- No ground truth or true labels
  - Cannot measure accuracy-based metrics
- Many different evaluation metrics
  - E.g. : Distortion score: average of the squared distances from the cluster centers of the respective clusters to each data point
  - Should not use them like we do in case of classification or regression
- Visualization is important
  - Application-dependent 'Meaning' is important

Finding Optimal K

- Application dependent and trial and error
- Check clustering metrics for a range of
  - E.g.: K = 2, 3, 4, 5, …, 10



- Elbow method
  - Choose a K such that adding another cluster doesn't lead to much better modeling of the data or the evaluation score
- Some models include finding K within the optimization procedure

Clustering: Challenges

- Correlated features
  - Demographic features such as family size and children at home may be correlated
  - Distance is calculated across more or less similar things

- High dimensionality of features
  - 'Distances' behave counter-intuitively in higher dimensions
  - Statistical learning in general not just clustering is difficult in higher dimensions
  - Let's get an intuitive understanding how clustering breaks down

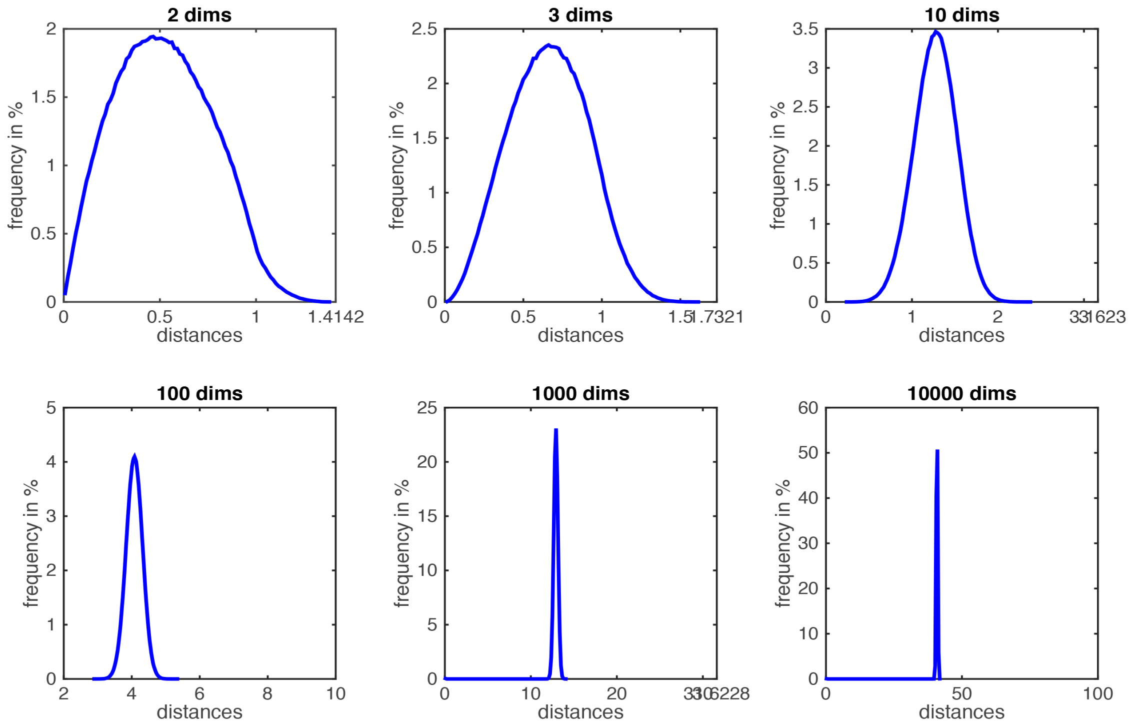
Curse of Dimensionality

- The clustering algorithms relies on the distance between two points
  - Makes the assumption that similar points share similar labels.

- In high dimensional spaces, points that are drawn from a probability distribution, tend to never be close together.

Curse of Dimensionality: Illustration

- Pair-wise distances between data point sampled *uniformly* in the d dimensional space



- At higher dimensions, all pairwise distances are concentrated at a higher value

http://www.cs.cornell.edu/courses/cs4780/2018fa/lectures/lecturenote02_kNN.html

Dimensionality Reduction

- Is it hopeless?
- Do all techniques break down?
  - No! Data often lies in a lower dimension than the total #variables (features)
- Dimensionality Reduction techniques
  - Principal Component Analysis
    - Reduces correlation
    - Transforms data into lower dimensional space

Principal Component Analysis

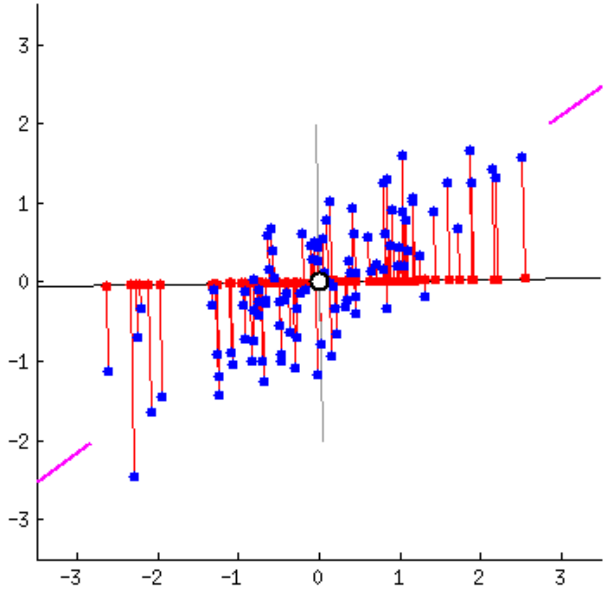
- Input: Data Matrix (rows: observations, columns: features)

1. *Center* each column
   - Subtract the column mean from each value in the column

2. Compute the *covariance matrix* among the features
   - Captures the collinearity among the features

Principal Component Analysis

3. Compute *eigenvectors* and *eigen values* of the covariance matrix: why?
   - Eigen vectors are oriented along the variance of the data
   - Eigen-values represent the magnitude
   - Covariance matrix is symmetric and its eigen vectors are always perpendicular to each other, so they have less correlation



Principal Component Analysis

4.1 Order the eigenvectors based on the value of their corresponding eigen-values to obtain principal components

- For p you will have p eigen vectors
- You may ignore less significant eigen vectors

4.2 Create a feature vector with selected eigen vectors and multiply it with the mean-adjusted data matrix to obtain transformed data

- *Note:* You may need to transpose for multiplying matrices

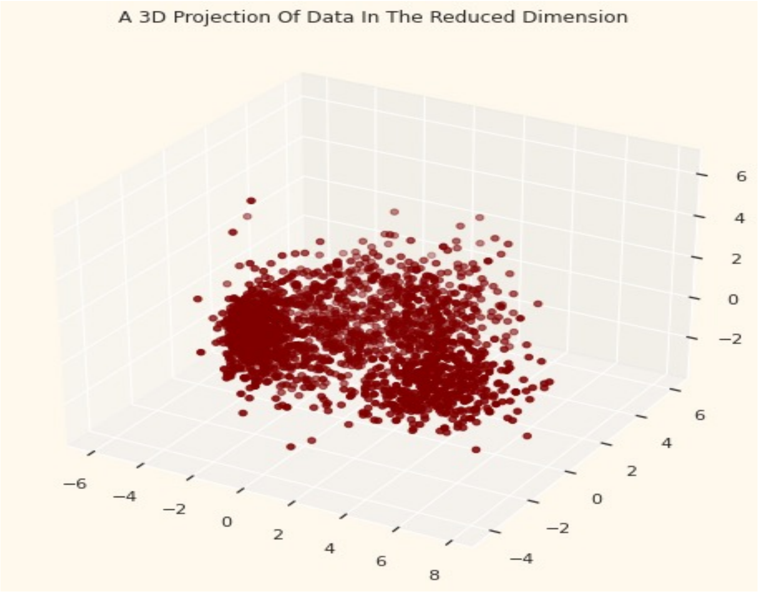
$$Feataure\ vector:\ [eig_1\ \ eig_2\ eig_3]$$

$$Final\ Data = Feataure\ vector \times Mean\ Adjusted\ Data$$

PCA: Disadvantages

- Difficult to interpret
  - Which is the key if you have to use it for targeting

- Visualization of the sample in reduced dimension:



Profiling

- Create clusters using PCA which represents the data in uncorrelated and low dimensional space

- Use raw variables to define the clusters
  - Summary statistics: mean, median, std. deviation, etc.
  - Visualization: Box plots or bar graphs

- Devise targeting strategies based on cluster characteristics

Profiling: Example

- Who are your your potential customers?
- Summary Statistics:

Clusters	Is_Parent	Family_Size	NumWebVisitsMonth	NumStorePurchases	Income	Teenhome	Kidhome	Total_Promos
0	1.000000	3.389262	5.817450	4.893960	47931.897987	1.012081	0.659060	0.136913
1	0.016162	1.595960	2.749495	8.305051	75251.709091	0.006061	0.010101	0.739394
2	0.769854	2.395462	6.818476	3.233387	30680.213938	0.056726	0.721232	0.092382
3	0.991549	2.656338	5.264789	8.709859	64913.625352	0.921127	0.101408	0.380282

Cluster 0: All parents with at least one teenager at home
Cluster 1: Single Households with high income
Cluster 2: Single parents with a kid
Cluster 3: Mostly parents with teenagers

Profiling: Example

- Where do you reach your potential customers?
- Summary Statistics:

Clusters	Is_Parent	Family_Size	NumWebVisitsMonth	NumStorePurchases	Income	Teenhome	Kidhome	Total_Promos
0	1.000000	3.389262	5.817450	4.893960	47931.897987	1.012081	0.659060	0.136913
1	0.016162	1.595960	2.749495	8.305051	75251.709091	0.006061	0.010101	0.739394
2	0.769854	2.395462	6.818476	3.233387	30680.213938	0.056726	0.721232	0.092382
3	0.991549	2.656338	5.264789	8.709859	64913.625352	0.921127	0.101408	0.380282

- For families with children, target them with online ads or online promotional campaigns
- For single households, provide deals at the store

Profiling: Example

- What products can you use to target your potential customers?
- Summary Statistics:

	Recency	Wines	Fruits	Meat	Fish	Sweets	Gold
Clusters							
0	48.284564	178.567785	8.524832	54.503356	11.604027	8.863087	25.641611
1	49.624242	607.402020	64.288889	464.844444	94.593939	65.014141	75.066667
2	48.320908	41.529984	6.662885	28.019449	10.077796	6.944895	19.380875
3	50.932394	608.380282	44.946479	229.518310	60.822535	47.202817	81.535211

- For single households (cluster 1), provide deals on wine products
  - One issue: data represents amount spent, if we can see quantity, it would be more useful

References

- Ch. 20 in Robert C. Blattberg, Byung-Do Kim and Scott A. Neslin, Database Marketing: Analyzing and Managing Customers

- http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf

Thank You