Data Mining

Neural Network and Linear Regression

https://data-mining.github.io/winter-2026/

CS 453/553 – Winter 2026
Yu Wang, Ph.D.
Assistant Professor
Computer Science
University of Oregon

Global Skeleton

1. **Overview and Mathematical Foundation (Linear Algebra, Probability/Statistics)**

2. **Classification**
    1. KNN
    2. Naïve Bayes
    3. Decision Tree

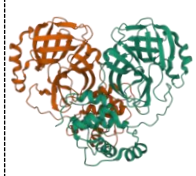
3. **Clustering**
    1. K-means Clustering
    2. Hierarchical Clustering
    3. KD-tree
    4. FAISS

4. **Dimension Reduction**
    1. PCA
    2. Linear Discriminative Analysis (LDA)

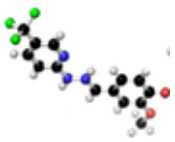
5. **Neural Network**
    1. Basic Architecture
    2. Linear Regression

Overview – What is Data?

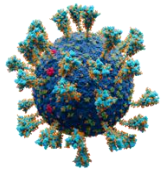
Science



Protein

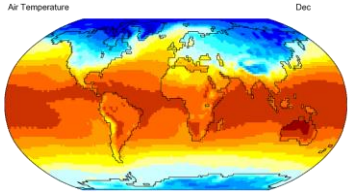


Small Molecule



Virus

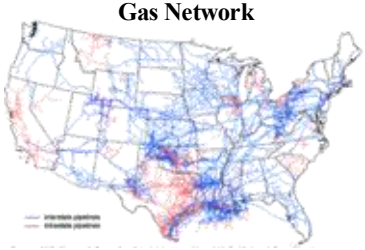
Brain Neural



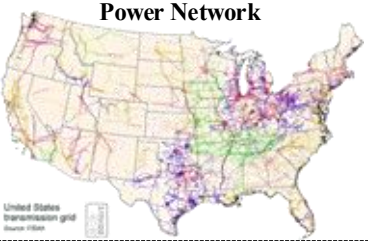
Surface Temperature of Earth

Infrastructure

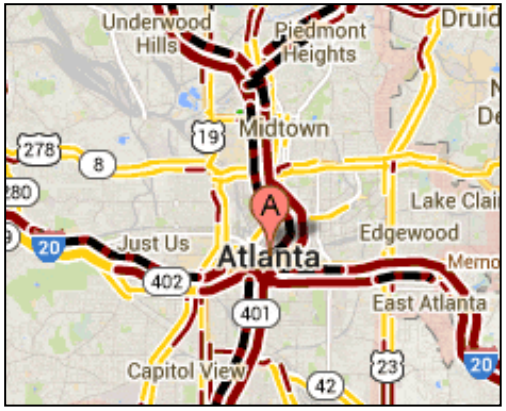
Gas Network



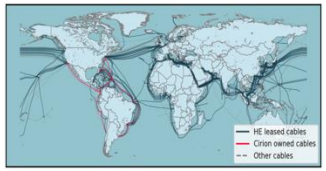
Power Network



Transportation Network

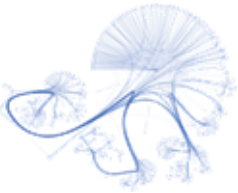


Submarine Cable



Terrestial Cable

Social Network



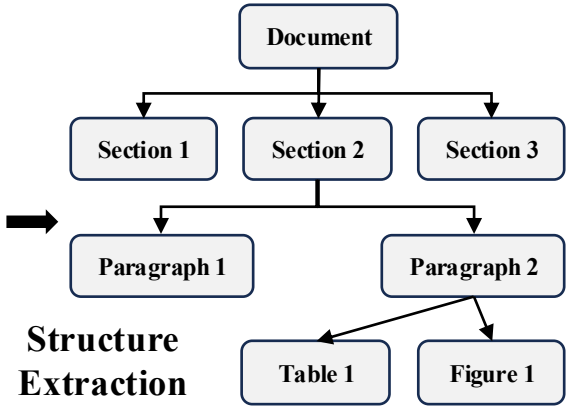
Citation Network

Transaction Network

User-Entity Interaction Graph

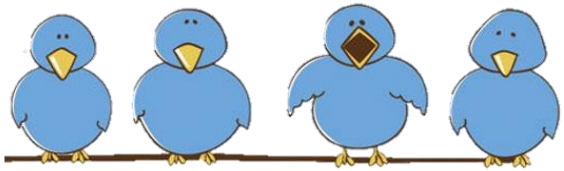
Virtual Village with AI Agents

Document

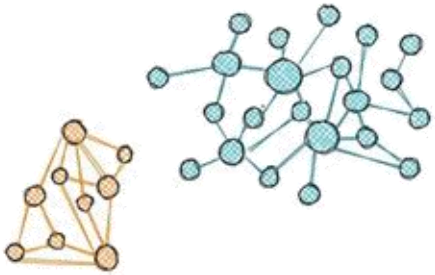


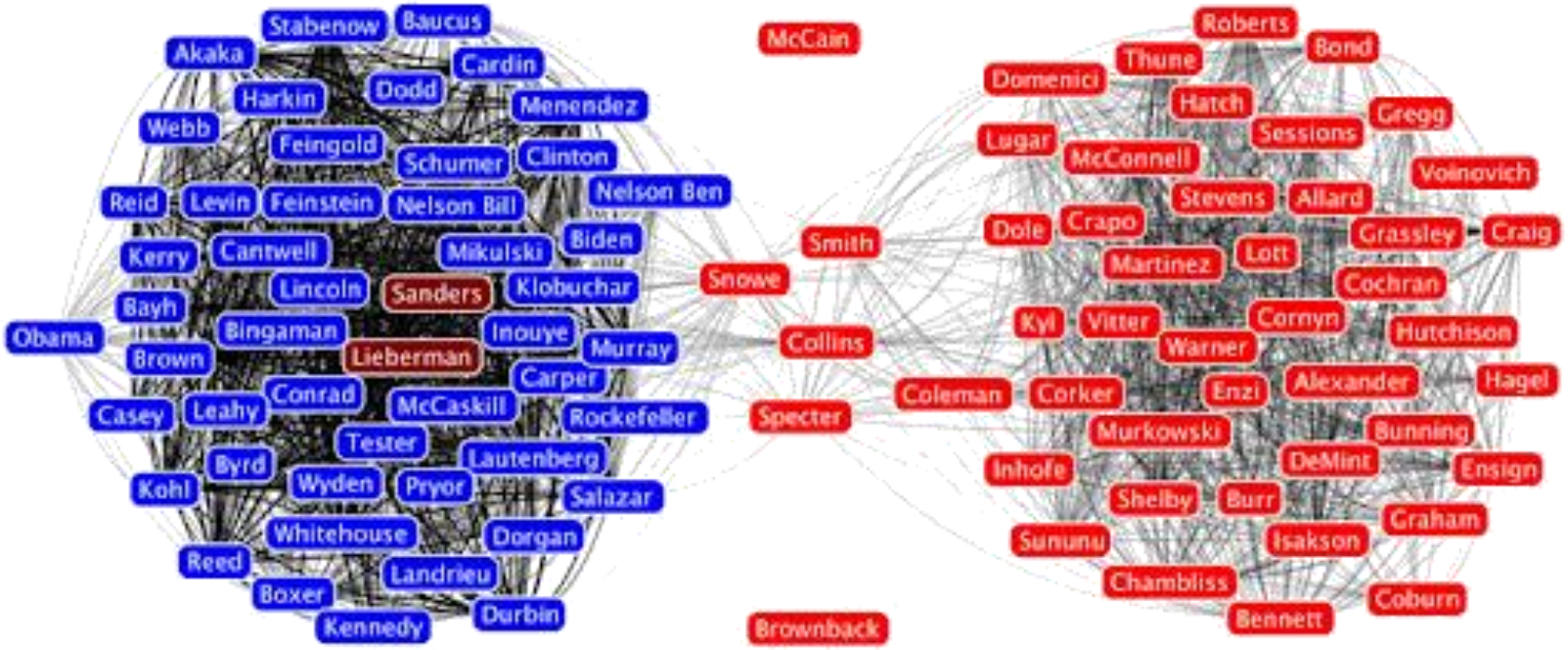
Structure Extraction

Overview – Why Analyze Data?



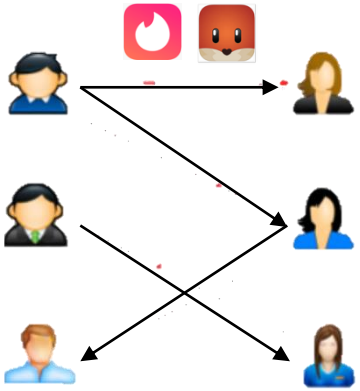
Birds of a feather flock together



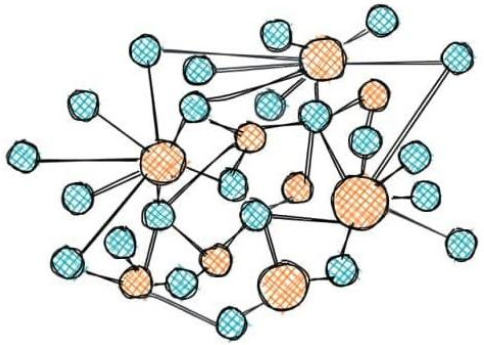


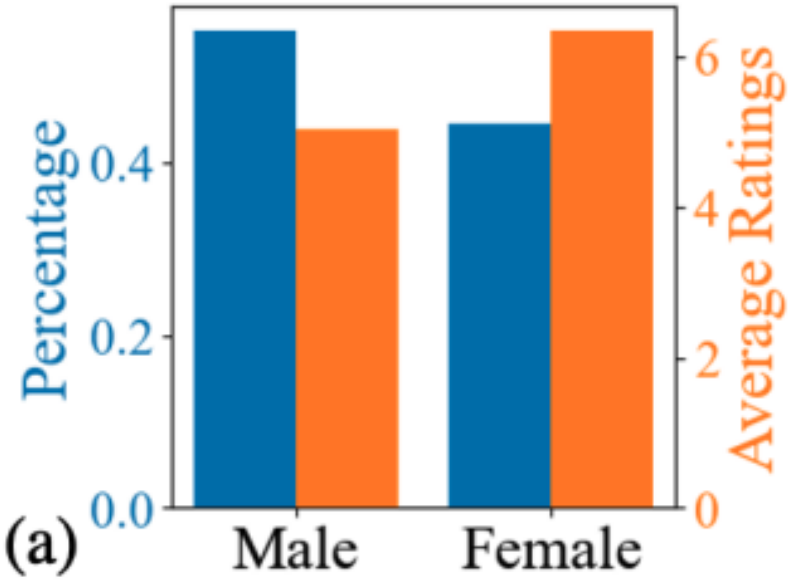
Gun Control Belief Network

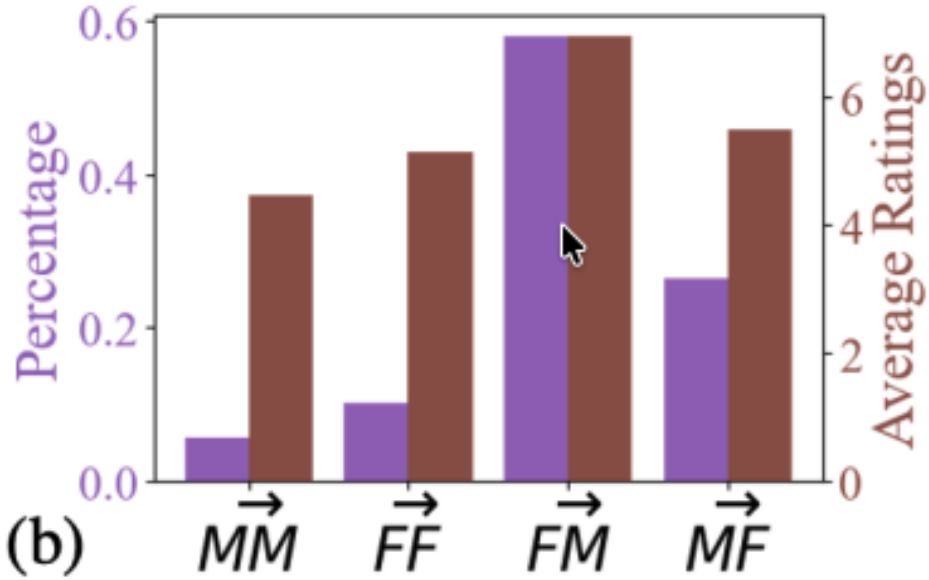
Overview – Why Analyze Data?



Dating Network

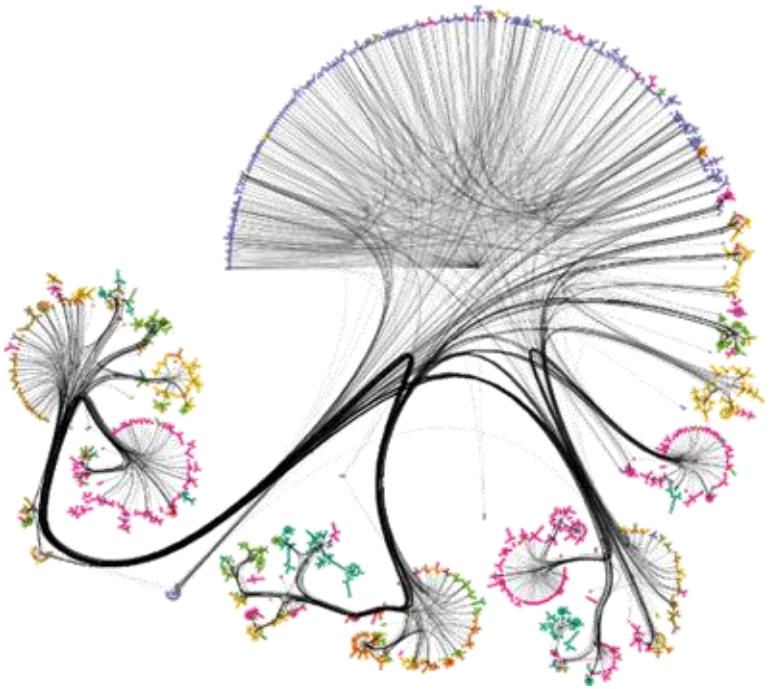






Dating Network

Overview – Why Analyze Data?



$$\frac{\sum_{e_{ij}\in\mathcal{E}} 1[y_i == y_j]}{|\mathcal{E}|}$$

$\mathcal{E}$ - Total Number of Edges

e_{ij} - Edge between node i/j

y_i - Label of i

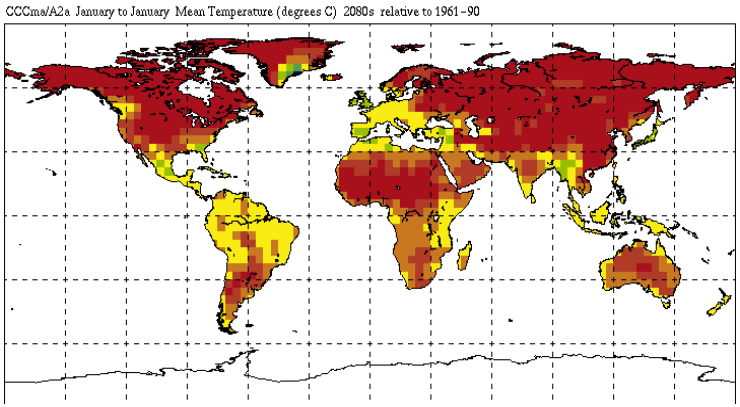
Table 5: Real data: mean accuracy ± stdev over different data splits. Best model per benchmark highlighted in gray. The "*" results are obtained from [26] and "N/A" denotes non-reported results.

	Texas	Wisconsin	Actor	Squirrel	Chameleon	Cornell	Cora Full	Citeseer	Pubmed	Cora
Hom. ratio h	**0.11**	**0.21**	**0.22**	**0.22**	**0.23**	**0.3**	**0.57**	**0.74**	**0.8**	**0.81**
#Nodes $\lvert\mathcal{V}\rvert$	183	251	7,600	5,201	2,277	183	19,793	3,327	19,717	2,708
#Edges $\lvert\mathcal{E}\rvert$	295	466	26,752	198,493	31,421	280	63,421	4,676	44,327	5,278
#Classes $\lvert\mathcal{Y}\rvert$	5	5	5	5	5	5	70	7	3	6

Overview – What are some challenges of mining data?

What kind of data mining question you want to answer?

Improving health care and reducing costs



Predicting the impact of climate change

Finding alternative/ green energy sources

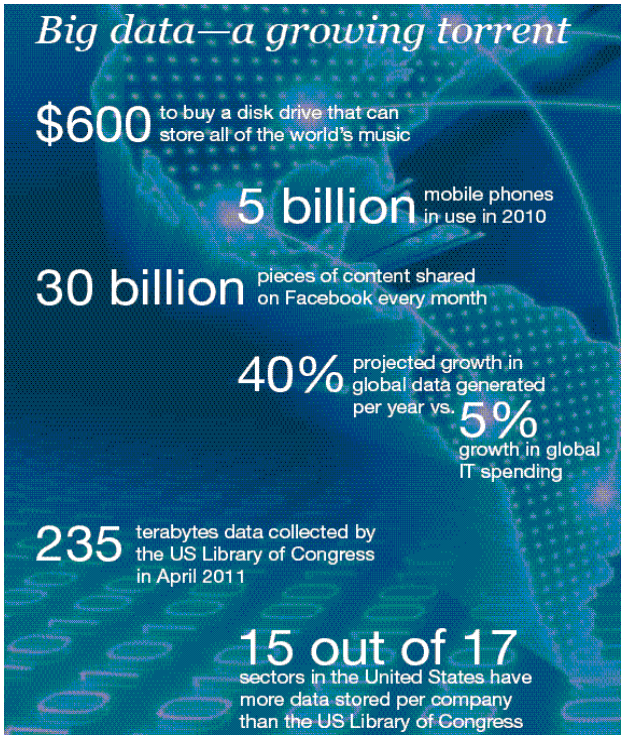
Reducing hunger and poverty by increasing agriculture production

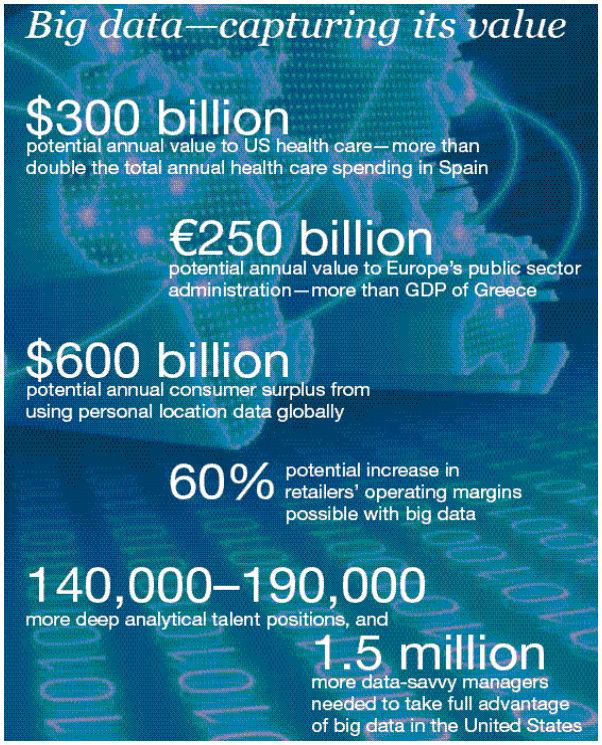
Overview – What are some challenges of mining data?

Data is usually in a very large scale!

McKinsey Global Institute

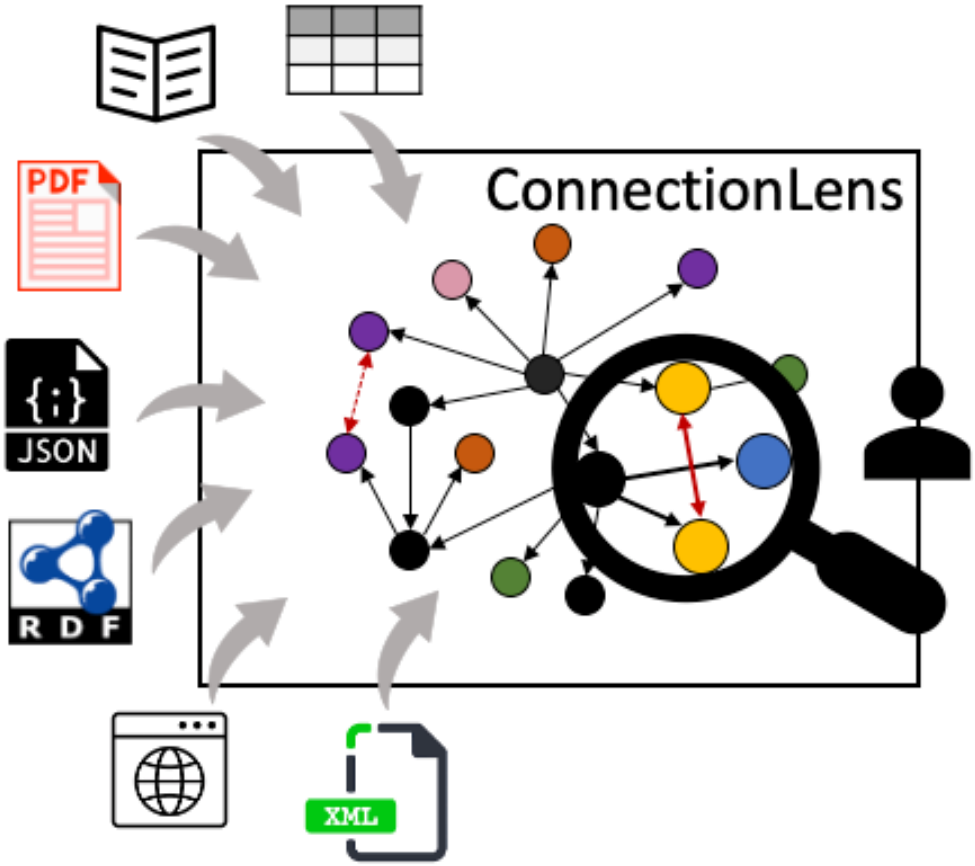
Big data: The next frontier for innovation, competition, and productivity





Overview – What are some challenges of mining data?

Data is diverse and heterogeneous



Stable Diffusion Image

Llama3 Language

Sora Video

Suno Audio

Scalar

$$\mathbf{v} = 3$$

Vector

$$\mathbf{v} = [1 \quad 2 \quad 5]$$

$$\mathbf{u} = \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}$$

Please note that we will use this one by default

Matrix

$$\mathbf{A} = \underbrace{\begin{bmatrix} 1 & 2 & 3 \\ 0 & 5 & 1 \\ 2 & 3 & 7 \\ 3 & 9 & 8 \end{bmatrix}}_{\text{3 columns}} \left.\vphantom{\begin{bmatrix} 1 & 2 & 3 \\ 0 & 5 & 1 \\ 2 & 3 & 7 \\ 3 & 9 & 8 \end{bmatrix}}\right\} \text{4 rows}$$

$$\mathbf{v} \in \mathbb{R}^{\mathbf{1\times 3}}$$

$$\mathbf{u} \in \mathbb{R}^{\mathbf{3\times 1}}$$

$$\mathbf{A} \in \mathbb{R}^{\mathbf{4\times 3}}$$

Matrix Multiplication

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 5 & 1 \\ 2 & 3 & 7 \\ 3 & 9 & 8 \end{bmatrix} \times \mathbf{B} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 5 & 7 \end{bmatrix} \rightarrow \mathbf{C} = \begin{bmatrix} 20 & 29 \\ 15 & 22 \\ 43 & 62 \\ \mathbf{61} & \end{bmatrix}$$

$3 \times 1 + 2 \times 9 + 5 \times 8$

4×3 $\quad$ 3×2

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 5 & 1 \\ 2 & 3 & 7 \\ 3 & 9 & 8 \end{bmatrix} \times \mathbf{B} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 5 & 7 \end{bmatrix} \rightarrow \mathbf{C} = \begin{bmatrix} 20 & 29 \\ 15 & 22 \\ 43 & 62 \\ 61 & \mathbf{89} \end{bmatrix}$$

$3 \times 2 + 3 \times 9 + 7 \times 8$

4×3 $\quad$ 3×2

Math Foundation – Linear Algebra

House 1
Size – 1000 sqft
2 bed, 2 bath
Location: 3

Contribution Coefficient 0.002, 1, 0.5, 1.2

$$\mathbf{X} = \begin{bmatrix} 1k & 2k & 1.5k \\ 2 & 3 & 2 \\ 2 & 2 & 3 \\ 3 & 2 & 4 \end{bmatrix}$$

House 2
Size – 2000 sqft
3 bed, 2 bath
Location: 2

$$\mathbf{A} = \begin{bmatrix} 0.002 \\ 1 \\ 0.5 \\ 1.2 \end{bmatrix}$$

House 3
Size – 1500 sqft
2 bed, 3 bath
Location: 4

$$\mathbf{Y} = \mathbf{A}^{\mathrm{T}}\mathbf{X}$$

Math Foundation – Distance

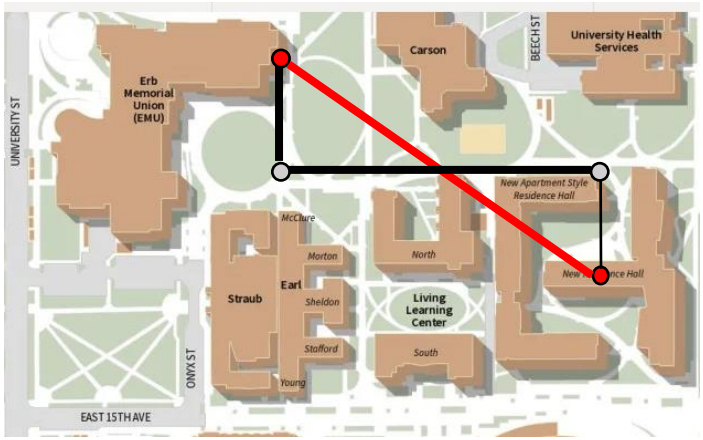
- $\|\mathbf{u} - \mathbf{v}\|^p$
- Function from a vector space to a single positive real value: $f: \mathbb{R}^d \rightarrow \mathbb{R}$
- Distance between **u** and **v**

$$\|\mathbf{u} - \mathbf{v}\|^p = \left(\sum_{i=1}^{d} |\mathbf{u}_i - \mathbf{v}_i|^p\right)^{\frac{1}{p}}$$

- Examples:

(1) Manhattan distance (L_1): $\|\mathbf{u} - \mathbf{v}\|^1 = \left(\sum_{i=1}^{d} |\boldsymbol{u}_i - \boldsymbol{v}_i|\right)$

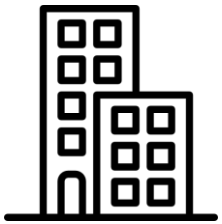
(2) Euclidean distance (L_2): $\|\mathbf{v}\|^2 = \left(\sum_{i=1}^{d} |\mathbf{u}_i - \mathbf{v}_i|^2\right)^{\frac{1}{2}}$



Math Foundation – Probability

P(B=W) = 0.3

P(ND|W) = 0.6



P(B=G) = 0.5

P(ND|G) = 0.2

P(B=S) = 0.2

P(ND|S) = 0.05

Math Foundation – Probability

P(B=W) = 0.3 P(ND|W) = 0.4

P(B=G) = 0.5 P(ND|G) = 0.8

P(B=S) = 0.2 P(ND|S) = 0.95

After one earthquake, the building is not collapsed

$$P(G|ND) = \frac{0.8 * 0.5}{0.71} = 0.56 \qquad P(S|ND) = \frac{0.95 * 0.2}{0.71} = 0.27$$

$$P(T|ND) = \frac{P(ND|T)P(T)}{P(ND)} \qquad P(W|ND) = \frac{P(ND|W)P(W)}{P(ND)} = \frac{0.4 * 0.3}{0.71} = 0.17$$

$$P(ND) = \sum_T P(ND|T)P(T)$$

$$P(ND) = \sum_T P(ND|T)P(T)$$

$$= 0.3 * 0.4 + 0.5 * 0.8 + 0.2 * 0.95 = 0.71$$

Math Foundation – Probability

P(B=W) = 0.17 P(ND|W) = 0.4

P(B=G) = 0.56 P(ND|G) = 0.8

P(B=S) = 0.27 P(ND|S) = 0.95

After two earthquake, the building is not collapsed

$$P(G|ND) = \frac{0.8 * 0.17}{0.77} = 0.177 \quad P(S|ND) = \frac{0.95 * 0.27}{0.77} = 0.33$$

$$P(T|ND) = \frac{P(ND|T)P(T)}{P(ND)} \qquad P(W|ND) = \frac{P(ND|W)P(W)}{P(ND)} = \frac{0.4 * 0.17}{0.77} = 0.09$$

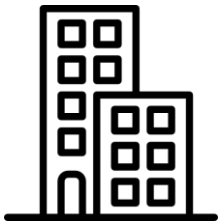
$$P(ND) = \sum_{T} P(ND|T)P(T)$$

$$P(ND) = \sum_{T} P(ND|T)P(T)$$

$$= 0.17 * 0.4 + 0.56 * 0.8 + 0.27 * 0.95 = 0.77$$

Math Foundation – Probability

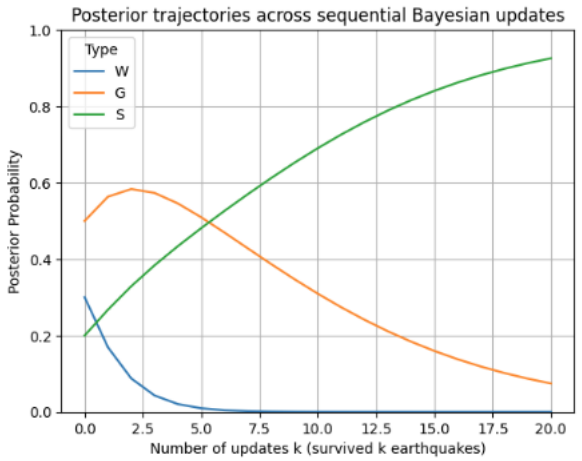
P(B=W) = 0.3

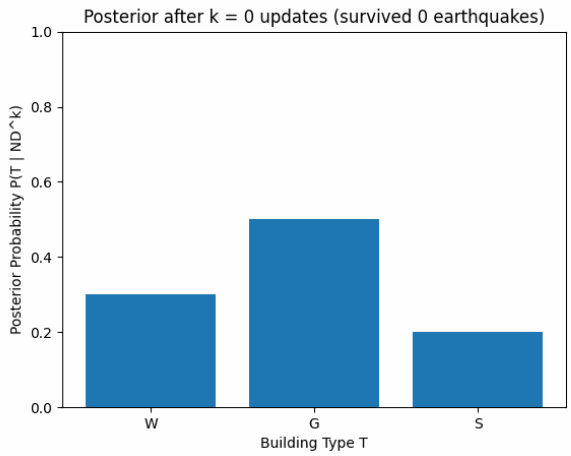


P(B=G) = 0.5

....

P(B=S) = 0.2





Math Foundation – Probability

Probability of the random variable X taking the value x $\quad P_X(X = x)$

$X \in \mathbb{R}^{256\times256}$: Image

x_1: Image1
5
$P_X(X = x_1)$

x_2: Image2
10
$P_X(X = x_2)$

x_3: Image3
5
$P_X(X = x_3)$

P_X

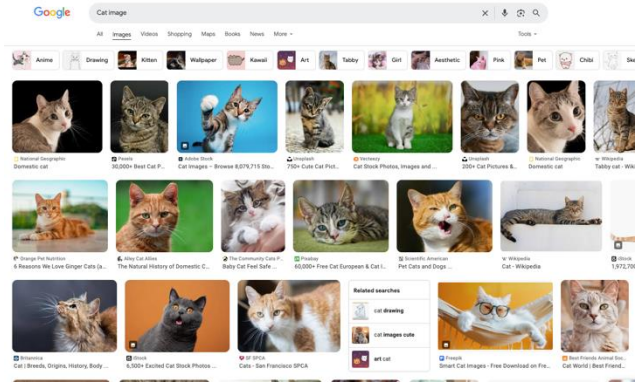
$P(X)$

Probability
Density
Distribution

Math Foundation – Probability

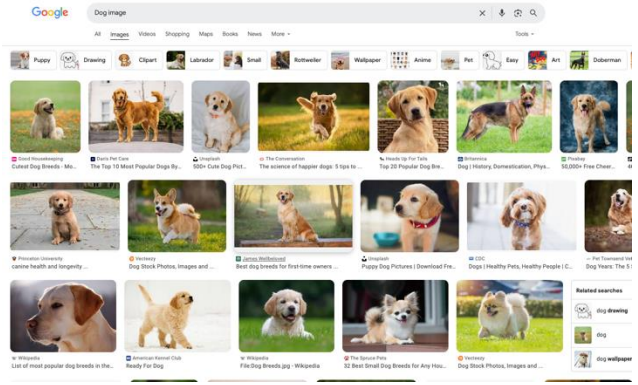
Probability of the random variable $X = x$ given $Y = y$ $P_{X|Y}(X = x|Y = y)$

Sampling something about



P(X|Y=Cat)

Sampling something about



P(X|Y=Dog)

Classification

1. **Naïve Bayes**
    1. **How to use Naïve Bayes to perform classification?**
    2. What is the main assumption of Naïve Bayes , how that is used in classification, and how that is a problem?

2. **Nearest Neighbor Classifier**
    1. **How to use KNN to perform classification?**
    2. Distance Choice – Document Example
    3. Recommender System Design
        1. Online – Offline
        2. Retrieval-Ranking
    4. Time Complexity
        1. Analyze Time Complexity
        2. Speed-up KD-Tree and Faiss

3. **Decision-Tree**
    1. **How to use Decision-Tree to perform classification?**
    2. How should split and when should we stop split?
    3. What is the problem of splitting too much or vice versa ?
    4. How to perform decision-tree classification on different types variables?
    5. What is the problem of decision-tree classifier?

Clustering

1. How to use K-means clustering to perform clustering?

2. Complexity Analysis

3. How to quantify the current performance of your clustering? Good or bad?

4. Convergence Analysis
    1. Will it always converge?
    2. Will it always converge to the same point?
    3. Will it always converge to the same good point?

5. Problem with K-Means Clustering
    1. K-Means Clustering with Feature Interaction

6. How to conduct Hierarchical Clustering?

7. FAISS

PCA

1. **What is PCA and why do we want to perform PCA?**

2. **Given a data point, how to project it on one perspective/one axis?**

3. **What is the objective function of PCA, how to mathematically express it?**

4. **What is the connection between PCA and eigen-decomposition?**

5. **Master PCA computational process and know how to use it.**
   1. Objective Formulation
   2. Relation to Eigen-Decomposition
   3. Use Derived Component to re-represent the data

6. **What is the difference between PCA and Linear Discriminative Analysis?**

Neural Network

1. **Neural Network and Biology Perceptron Connection**

2. **The very basic and most fundamental neural network format (Diagram and Formula)**
    1. Linear Transformation
    2. Bias
    3. Nonlinear Attention

3. **How to use Neural Network to perform Linear Regression?**
    1. Feature Transformation for prediction
    2. MSE loss to represent objective
    3. Optimization
        1. Theoretical Optimization
        2. Computational Optimization – Gradient Descent

Question Time!