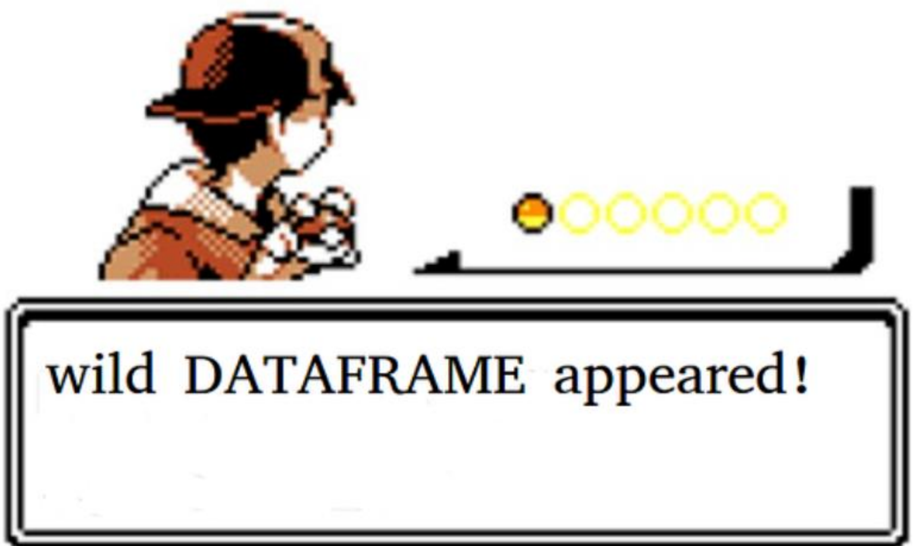
CISC 372
T2 Review

	name	age	state	num_children	num_pets
0	john	23	iowa	2	0
1	mary	78	dc	2	4
2	peter	22	california	0	0
3	jeff	19	texas	1	5
4	bill	45	washington	2	0
5	lisa	33	dc	1	0

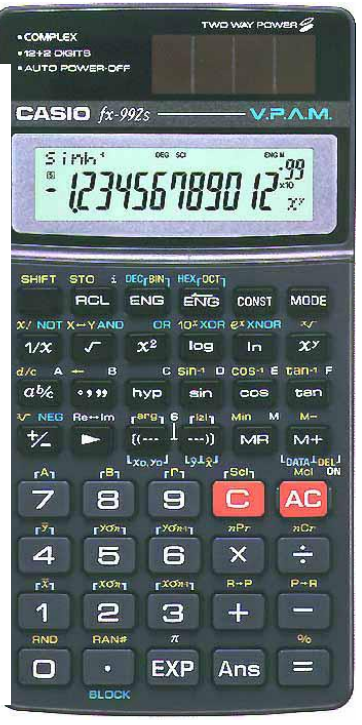


IT IS REWIND TIME

https://www.youtube.com/watch?v=YbJOTdZBX1g

Calculator with “log” function

- The following models are recommended.
- CASIO
  - fx-100MS, fx-115MS,
  - fx-260, fx-570MS,
  - fx-991MS, fx-992S
- SHARP
  - EL-510, EL-520,
  - EL-531, EL-546
  - Models extensions are acceptab



- We don't need calculator any more in the second test
- But you can use one if you want

Test #1

- Please return the Quiz
  - If you want to keep you can take pictures
  - I need to keep the HARD copy for two years.. 
  - *I won't enter your grade if you don't return the copy..*

Association Rule Mining

- Apriori Algorithm

- Confidence vs. Support vs. Lift (interestingness)

NB – Naïve Bayesian

- Decision Boundary
- Generative model
  - Can be either Parametric or non-parametric
  - Depends on how one models the class conditional probability
- Advantage:
  - Interpretable prediction
  - In most cases work well with small dataset
- Disadvantage:
  - Assume variable independence

Naïve Bayesian Classifier

$P(C_i)$: P(buys_computer = “yes”) = 9/14 = 0.643

P(buys_computer = “no”) = 5/14= 0.357

X = (age <= 30 , income = medium, student = yes, credit_rating = fair)

Compute **$P(X|C_i)$** for each class

P(age = “<=30” | buys_computer = “yes”) = 2/9 = 0.222

P(age = “<= 30” | buys_computer = “no”) = 3/5 = 0.6

P(income = “medium” | buys_computer = “yes”) = 4/9 = 0.444

P(income = “medium” | buys_computer = “no”) = 2/5 = 0.4

P(student = “yes” | buys_computer = “yes) = 6/9 = 0.667

P(student = “yes” | buys_computer = “no”) = 1/5 = 0.2

P(credit_rating = “fair” | buys_computer = “yes”) = 6/9 = 0.667

P(credit_rating = “fair” | buys_computer = “no”) = 2/5 = 0.4

age	income	student	credit_rating	comp
<=30	high	no	fair	no
<=30	high	no	excellent	no
31…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

$P(X|C_i)$: P(X|buys_computer = “yes”) = 0.222 x 0.444 x 0.667 x 0.667 = 0.044

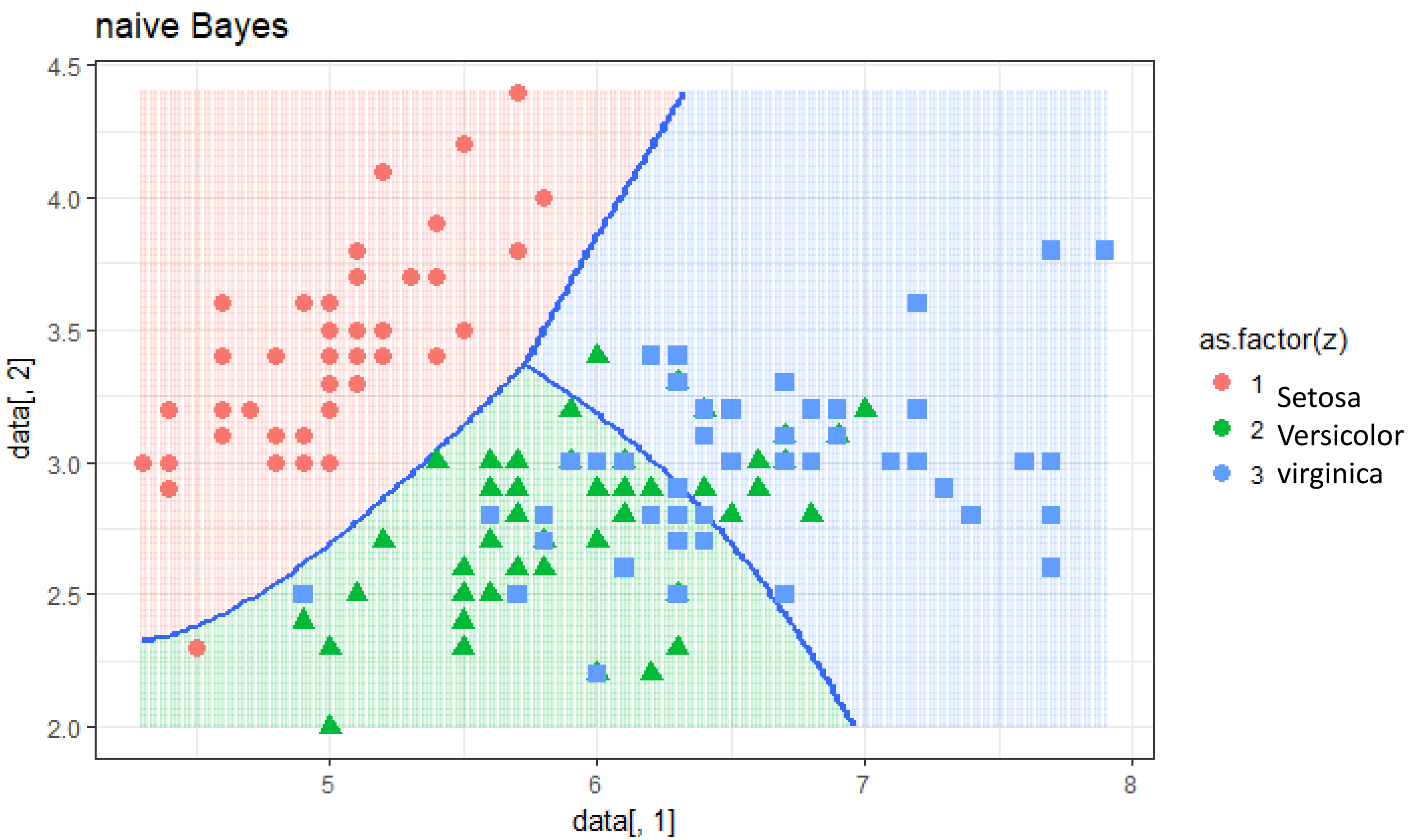
P(X|buys_computer = “no”) = 0.6 x 0.4 x 0.2 x 0.4 = 0.019

$P(X|C_i)*P(C_i)$: P(X|buys_computer = “yes”) x P(buys_computer = “yes”) = 0.028

P(X|buys_computer = “no”) x P(buys_computer = “no”) = 0.007

Therefore, X belongs to class (“buys_computer = yes”)

Naïve Bayesian – Decision Boundary



NB – Naïve Bayesian

- Decision Boundary
- Generative model
  - Can be either Parametric or non-parametric
  - Depends on how one models the class conditional probability
- Calculation

Clustering

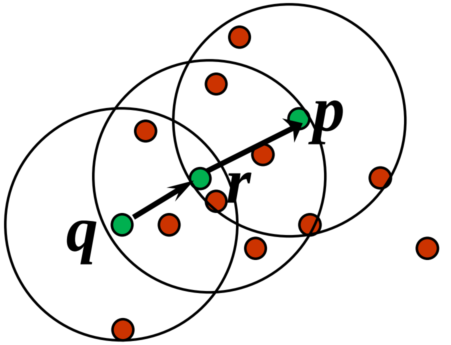
- Partitioning Methods vs Hierarchical Methods
  - Difference
  - K-mean/K-Medoids/AGNES/DIANA
- K-mean vs K-Medoids (PAM)
  - Outlier?
  - Complexity?
- Agglomerative vs Divisive method
- Density-based Clustering

Clustering

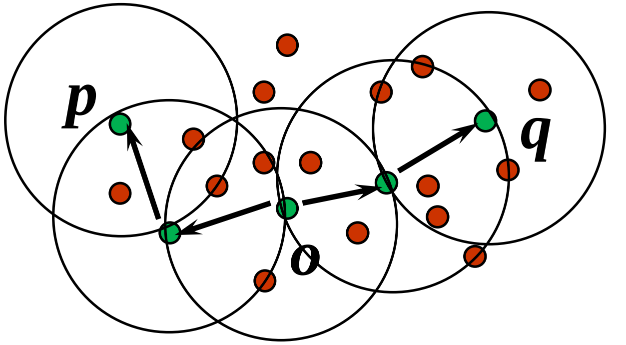
- Partitioning Methods vs Hierarchical Methods
  - Difference
  - K-mean/K-Medoids/AGNES/DIANA
- K-mean vs K-Medoids (PAM)
  - Outlier?
  - Complexity?
- Agglomerative vs Divisive method

Density-based Clustering

- Two parameters*:*
  - ε: Maximum radius of the neighbourhood
  - *MinPts*: Minimum number of points in an ε-neighbourhood of that point
- $N_\varepsilon(q)$: $\{p \mid dist(p,q) <= \varepsilon\}$



- *Concepts:*
  - *Core point*
  - *Directly density reachable*
  - *Density reachable*
  - *Density-connected*



DBSCAN

- Can detect outliers
- Can detect arbitrary shape clusters
- Does not need to know the number of clusters
- Resistant to noise
- Efficient (a single pass over the data points)
- Problem: need to define density

Tuning

- Three typical ways:
  - Grid search – global optimization
  - Random search – local optimal
  - Bayesian optimization – local optimal

Text Analytics

- Understand the problem:
  - Many-to-one
  - Many-to-Many
- Preprocessing
  - Stemming
  - Case normalization
  - Stop words & Punctuation removal

Text Analytics

- Models
  - BOW
  - N-gram model
  - character n-gram model
  - N-gram vs. n-perm

- Representation
  - Term frequency (TF)
  - Term_frequency / document_frequency (TF-IDF)

RNN

- Compared to n-gram, why RNN??
- Issues in RNN:
  - Long dependency
  - Gradient Vanishing/Explosion
- Cell implementation:
  - Vanilla vs. GRU vs. Attention vs. Multi-head Attention
  - Difference in design and ***why***

Language Model

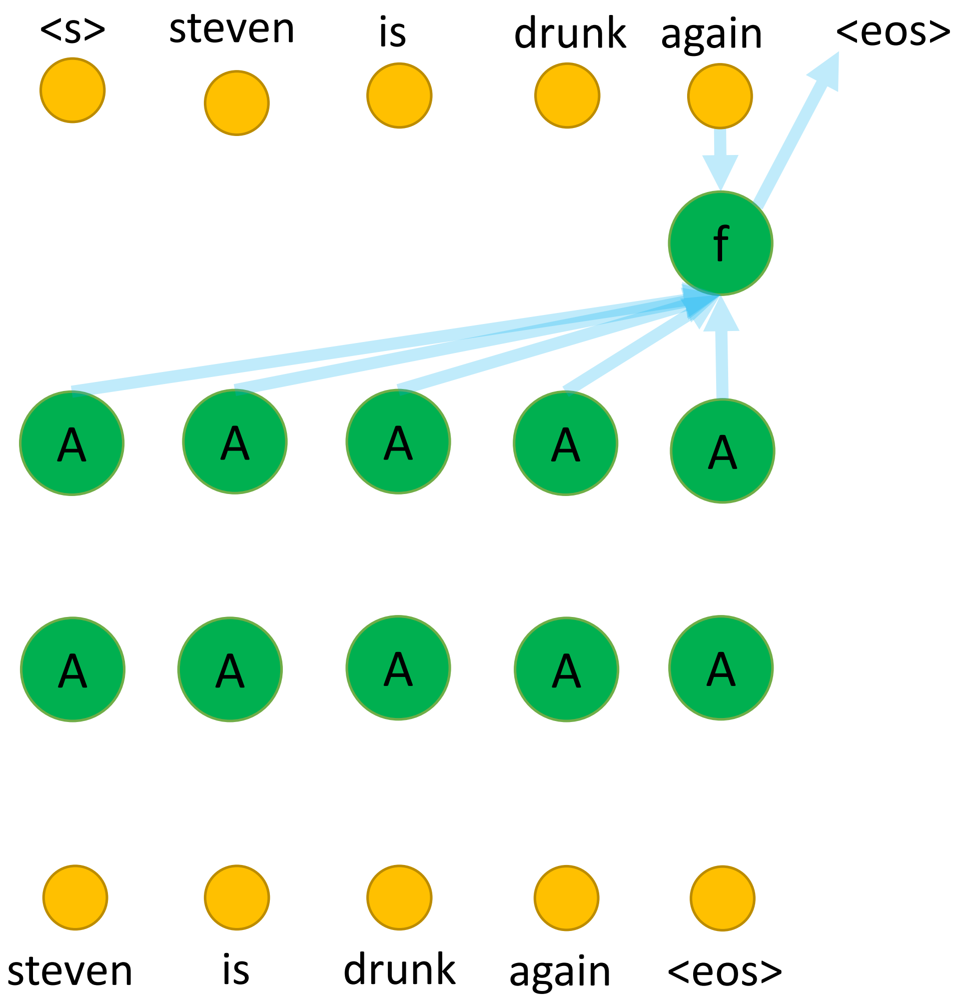
- Why we need language model
  - Foundation of various down-stream tasks
    - (translation etc.)
  - Foundation of representation learning
  - Foundation of semi-supervised learning

- N-gram modeling
  - The ***longer*** the context, the more coherent
  - Problem?

- **Word2Vec: CBOW vs Skipgram (design difference)**

What is Transformer

The Transformer – Attention is all you need



http://karpathy.github.io/2015/05/21/rnn-effectiveness/