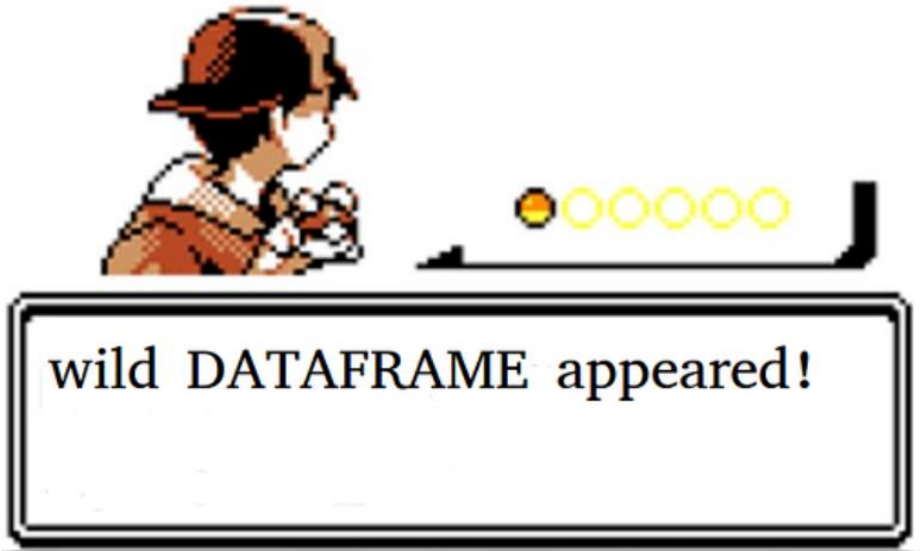
CISC 372
T1 Review

	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

E-mail Policy

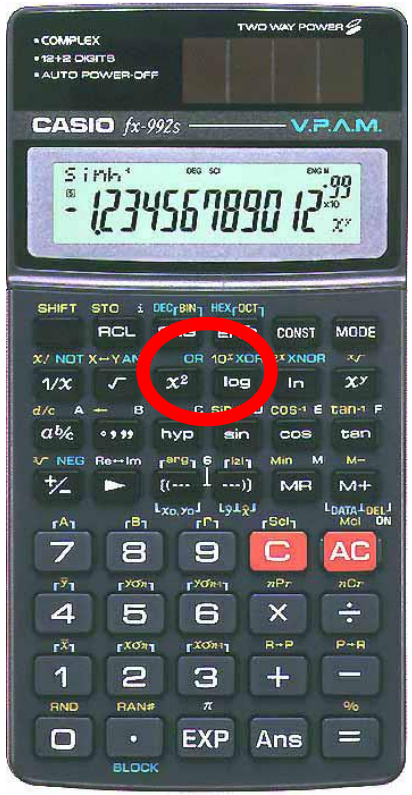
- When you send e-mail to me, put **"CISC 372"** in the subject area, so that it can pass the spam filter.

- Visit me during my office hour if questions required extensive explanation. (or if you want to chat with me on something you are interested)

- **Course email list is a must-read**.

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 acceptable.



Background (L1, L3)

- Data Science – A new approach to understand/model unknown system
  - Empirical science, theoretical science, computational science

- Security:
  - **CIA** for security evaluation
  - Security/Ethical implication of AI/DS

Experimental Protocol (L4)

- Supervised vs. Unsupervised Learning (& semi-)
- Classification vs. Prediction/Regression
- Macro-average vs. Micro-average
- Bias vs. variance
- Overfitting vs. underfitting

- Hold-out method (with validation set)
  - Purpose of different set
- Cross-validation
- .632 Boostrapping

- Hyperparameter tuning (as a DS life cycle)

Understanding the data (L5)

- Data attribute
  - Ordinal vs. categorical

- Preprocessing
- Normalization
- Standardization
- Discretization

Classifiers/Regressors (L2, L7-L11)

- Know decision boundaries (what it may look like)
- What are the limitations?
- How to regularize (if any)?
- Generative model/Discriminative model?
- Eager learner/Lazy learner?
- Parametric model/Non-parametric model?
- Flexible/Inflexible (w.r.t. different hyperparameter)?
  - So you know what to do to reduce bias
  - You know what to do to reduce variance
- Interpretability

Decision Tree[s] (for classification) L7-L8

- How to create a node in a tree, given a dataset.
  - ***Calculator to calculate log with base 2***
- What is the problem of information gain?
- How **gain ratio** solve this problem?
- What are the metric used for selecting features?
  - ID3
  - CART
  - C4.5

- How to handle numeric value attribute
- Pre-pruning/Post-pruning
- Random Forest vs Single Tree
  - Build-in bootstrapping

XGboost L9

- Objective function?
  - Loss function + regularization
- Difference compared to random forest?
  - Optimize tree selection toward an objective function
- How does it regularize?
- How to adjust its flexibility?

Neural Network (L10-L11)

- CNN vs NN
  - Improvements
- Pooling Layer vs Convolution Layer
- GD
  - Stochastic GD (SGD) (1-sample/mini-batch)
  - Momentum
  - Adagrad:
    - Adaptive learning rate
    - Issues? Methods proposed to solve this issue.

Instance-based Learning & Naïve Bayesian L11

- K-NN
  - What are the decisions one needs to make?
  - How to adjust flexibility

- Naïve Bayesian
  - Generative model
  - Easy to implement
  - Require a smaller set of training data
  - Issue: class conditional independence

What is ***not*** covered?

- Clustering
  - Partition-based clustering
    - No k-mean
    - No k-medoids
  - Density-based clustering

- Feature Selection
- Handling Missing Values
- Naïve Bayesian
- Semi-automated hyperparameter tuning