

Final exam topics

CS251: Data analysis and visualization

Spring 2019

Note: I am providing this guide to help out with studying, I make no claim that this list of topics is exhaustive. A good study tool is to review/redo homework and quizzes.

What's not on the exam

- Tkinter and GUI-specific concepts
- Interactive viewing control laws

Visualization

- Terminology
- Components of a plot
- Types of visualization
- Good vs. bad visualizations
- How data visualization can be misleading, factors to consider with human visual perception
- 2D coordinate transformations (getting data to appear in a plot)
- 3D coordinate transformations (getting data to appear in a plot)
- Translation matrix
- Scaling matrix
- Rotation matrix
- Orthographic viewing pipeline
- Histograms and binning
- Kernel density estimates
- Scatterplot

Analysis

Computational tools

- Numpy (vectorization, broadcasting, dimensions/axes)

Probability

- Probability basics
- Binomial distribution
- Gaussian distribution
- Bayes Rule
- Likelihood, prior, posterior distributions
- Bayesian updating
- Covariance matrix

Linear regression

- Simple linear regression
- Basis functions, coefficients, etc
- R^2
- Normal equations
- Multiple linear regression

Principle component analysis

- Covariance matrix approach
- Singular value decomposition approach
- Data reconstruction

Clustering

- Distance metrics
- The clustering problem
- K-means
- Leader algorithm
- Hierarchical clustering
- Single-linkage algorithm
- Elbow method
- Number of clusters

Machine learning

Classification

- Supervised learning
- Binary classification problem

- K Nearest Neighbors
- Initialization
- Performance considerations
- Naive Bayes: One feature and multiple features

Binary classification

- Confusion matrix
- Binary classification metrics
- ROC Curve
- Percent correct, AUC/C-Index

Theorems

- No Free Lunch Theorem
- Ugly Duckling Theorem

Decision trees

- Entropy
- Information gain
- ID3 algorithm
- Pruning
- Overfitting
- 1R decision trees

Neural networks

- McCulloch & Pitts Neurons
- Single layer neural networks
- ADALINE
- Perceptron
- Delta rule(s), error-based learning