

note

127 views

Final topics

The exam will be on December 12th at 8:30am in HEBB 100, and it will be for the standard amount of time (2 hours and 30 minutes, I believe).

It will be closed book, but you can bring in 4 pieces of paper with whatever you want written on either side of them.

The following are topics that may be on the final:

Assignment 1:

Summary statistics (like range, median, and quantiles).

Data visualization (histogram, scatterplot, and box plot).

Decision trees (how to fit stumps using classification error, how to classify a new example, runtime in $O()$ notation, effect of depth).

Learning theory (training vs. test error, validation sets and cross-validation, fundamental trade-off)

Naive Bayes (conditional probability, fitting the probabilities, classifying a new example, runtime).

Assignment 2:

K-nearest neighbours (how to classify a new example, runtime in $O()$ notation, effect of k , condensed version).

Random forests (what bootstrap does, how random trees works, effect of number of trees/random-features, how to classify new example).

K-Means (effect of initialization, error functions, runtime, how to cluster new example, elbow method, k-medians, vector quantization).

Density-based clustering (effect of parameters, shape of clusters).

Assignment 3:

Vectors/matrices/norms (summation and vector/matrix/norm notation, minimizing quadratic functions as linear systems)

Linear regression (adding bias, change of basis, RBFs, regularization, cost of training/testing, effect of basis parameters).

Robust regression (weighted least squares, smooth approximations, computing gradients).

Assignment 4:

Regularized logistic regression (loss functions for binary classification, effect of different regularizers on overfitting and sparsity)

Convex and MLE/MAP estimation (showing functions are convex, connection between probabilities and losses/regularizers)

Multi-class Logistic (one-vs-all, softmax loss derivatives and implementation)

Assignment 5:

PCA (computing 1st PC, scaling issue, PCA for visualization, PCA for compression)

Beyond PCA (different loss functions, effect of regularizers)

Multi-dimensional scaling (basic model, ISOMAP and geodesic distance)

Assignment 6:

PageRank (random walk model)

Stochastic gradient and neural networks (effect of step-size, effect of standardization, effect of initialization, effect of non-linearity, effect of regularization)

To give you an idea of what to expect for the final, I will update this post at the end of the week with the final and practice final from last year.

#pin

exam

Updated 4 days ago by Mark Schmidt

followup discussions *for lingering questions and comments*