6.036 Machine learning

Review lecture

Final Exam (9:00am, May 25)

- 3h exam, i.e., more comprehensive than midterm
- closed book, you cannot bring a cheatsheet
- we will give you a cheatsheet at the exam (and post it on-line ahead of time, at least two days before)
- it covers the whole course material but has roughly speaking 2/3 emphasis on the material since midterm
- questions are of the same kind as in the midterm
- there can be broader questions from application lectures

- previous finals (as practice problems) already available on Stellar (with solutions)

Review sessions

- Tomorrow (Friday, May 19) during recitation times in 26-100 (here)
  - 11-2 guided review
  - 2-5 office hour style Q/A
- We will also offer additional office hours during the finals week (see the corresponding piazza announcement)

Material before the midterm (1/3)

- Linear classification, separation
- On-line linear classifiers
- Maximum margin separation; SVM, Pegasos
- Linear regression
- Collaborative filtering
- Non-linear classifiers, kernels
- Neural networks, deep learning, back-propagation
- Recurrent neural networks

Material since the midterm (2/3)

- Generalization, VC-dimension
- Clustering (k-means)
- Mixture models
- EM algorithm
- Hidden Markov Models
- Bayesian Networks
- Reinforcement Learning

- Also part of the course material:
  - applications: computer graphics/vision
  - applications: natural language processing
  - applications: machine learning + hardware

Generalization, VC-dim

- Need to understand (qualitatively) how the difference between training and test errors depend on the number of training examples (n) and the (log of the) number of classifiers considered (log|H|)
- Shattering and VC dimension, calculating the VC dimension of a simple set of classifiers

- Note that there are many ways to understand generalization: how log-likelihoods of training and test data for a mixture model behave as we change the number of mixture components

Clustering

- K-means clustering
- How the algorithm behaves
  - with a different initialization
  - with a different choice of metric
- What K-means fails to capture

Mixture models

- Definition of a mixture of spherical Gaussians, what the parameters are, what they control
- How to sample from a mixture model
- Types of data that the mixture model is designed to model well (e.g., overlapping clusters)

- We use the (log-)likelihood of the data to measure how well a mixture model agrees with the data. Need to understand (qualitatively) how this log-likelihood behaves if we change the mixture parameters

EM algorithm for mixtures

- E-step of the EM algorithm (cf. assignment step in the k-means algorithm)
- M-step of the EM algorithm (cf. update of cluster centroids in k-means)

- Should be able to assess by hand how the algorithm modifies the mixture in simple cases (cf. previous exam questions)
- Need to understand the effect of different initializations on the resulting model
  - e.g., if all the mixture components are initially the same
  - e.g., same means and mixing proportions, different variances

Hidden Markov Models

- Need to understand them "as models" (cf. previous exam questions, example at lecture)
- If given the initial state distribution, transition probability matrix, and the emission (output) distribution, you should be able to evaluate
  - the probability of a particular sequence of states and observations
  - the probability of a sequence of observations (not knowing the states)
  - the most likely sequence of hidden states corresponding to a sequence of observations
  - the probability that the state at time t took a particular value if given the observations

Bayesian networks

- Understand "arcs" as dependencies, interpret multiple incoming "arcs"
- The notion of probability tables (e.g., one variable depending on two others)
- How to write the distribution over all the variables as a product of simpler terms
- Basic independence statements

- Be able to draw Mixture models, Markov, and Hidden Markov Models, etc. as Bayesian networks

MDPs, reinforcement learning

- Definition in terms of states, actions, rewards, transition probabilities, discount factor
- What value function is, how to calculate it through value iteration (or from Q-values)
- Q-values (what they mean), how to get them from Q-value iteration (or based on the value function)
- policy derived from Q-values; optimal policy

- For example: should be able to calculate how the policy derived from Q-values changes as a function of Q-value iterations (cf. previous exam questions)

Applications

- Broader questions about the topics
  - different from last year
  - for level, check the previous exam questions

Good Luck!

- Some tips
  - the final exam is cumulative so don’t forget the midterm material
  - most questions pertain to topics well-covered in the course (homeworks, projects, etc.)