
CAP 5636 – Final

Date: Thursday, December 7, 2017 – 4:00 – 7:00pm

Name:

Instructions:

- This exam is open book and open notes. Textbooks and notes on tablet devices are acceptable but they must be put into airplane mode. No device with a keyboard is acceptable.
- It is recommended that you use a pencil, such that you can make corrections. Do not use highlighters, and don't use red colored pens.
- Allotted time is 180 minutes.
- Note that the points add up to 100 + 20 bonus points.

Problem 1 (Search - 30 pts)

a) Explain the difference between tree search and graph search. (2 sentences)

In tree search a state represents an entire path in a search state graph, the same state might be visited multiple times if reached by different paths. In graph search each state is visited and expanded only once.

b) Give one advantage and one disadvantage of graph search over tree search.

+ As in graph search each state is expanded only once, graph search may be faster, as repeated expansions are avoided.

-Additional memory is needed to keep the set of expanded states.

-Expansion strategies that yield optimal result with tree search might not give an optimal result with graph search, as better solutions will not be expanded.

Problem 2 (Classifiers - 30 pts)

a) Explain the difference between model based and not-model based classifiers

Model based classifiers build a probabilistic model of the phenomena, instantiate the known variables, and query the model for the probability of the classification. Examples include Naïve Bayes models, and any other Bayes Nets. Classifiers that are not model based include decision trees, perceptrons, other neural networks, K-nearest neighbor etc.

b) It is recommended that for machine learning, one separates label data into training data, held out data and test data. Explain the need of held out data.

The training data is the data set on which the learning algorithm is running. The algorithm should only see the test data once. Thus we can use the held-out data to choose or tune the learning algorithm - in many cases this amounts to adjusting the hyperparameters such as NN structure, the k of Laplace smoothing and so on.

Problem 3 (Particle Filters - 30 pts)

The particle filter cycle consists of the steps of 1) elapse time, 2) observe and 3) resample. Let us consider that we are using a particle filter to track the location of a robot 3 steps into the future. We choose to not do the resampling step.

a) Give two advantages we might gain by doing this.

1.

We don't spend time to do the resampling.

2.

The resampling might probabilistically lose information.

b) Give two disadvantages we might incur by doing this.

1.

We lose performance if we spend our computation time on particles with low weights.

2.

We lose accuracy if we spend our computation time on particles with low weights.

Problem 4 (Naïve Bayes - 30 pts)

Let us assume that our training data contains the following messages:

we have an exam in AI (HAM)
we are going to cinema (HAM)
you should buy this flashlight (SPAM)

Estimate using both maximum likelihood and Laplace smoothing with $k=1$. Assume that there are no other words possible except those listed, and assume that the size of the outcomes is the number of words.

$P(\text{should} \mid \text{HAM})$

Words: 1 we 2 have 3 an 4 exam 5 in 6 AI 7 are 8 going 9 to 10 cinema
11 you 12 should 13 buy 14 this 15 flashlight

$$P_{\text{MLE}}(\text{should} \mid \text{HAM}) = 0 / 2 = 0$$

$$P_{\text{L1}}(\text{should} \mid \text{HAM}) = (1) / (2 + 1 \cdot 15) = 1 / 17$$

$P(\text{should} \mid \text{SPAM})$

$$P_{\text{MLE}}(\text{should} \mid \text{SPAM}) = 1 / 1 = 1$$

$$P_{\text{L1}}(\text{should} \mid \text{SPAM}) = 2 / (1 + 15) = 2 / 16 =$$

Classify using Naïve Bayes the following email. Estimate the probabilities using Laplace smoothing with $k=1$

should have an flashlight.

$$P_{\text{HAM}} = 1/17 * 2/17 * 2/17 * 1/17 = 4 / (17^4) = 4.78E-5$$

$$P_{\text{SPAM}} = 2/16 * 1/16 * 1/16 * 2/16 = 4 / (16^4) =$$

Classified as SPAM.