

Assignment 3– EM and Bayes Nets.

Assignment overview. This assignment is designed to show some examples of Bayesian Networks and generative models. This assignment requires you to apply a Gaussian mixture model to the Iris dataset with the Expectation Maximization (EM) algorithm. The second question is an example to apply a probabilistic modeling package called Lea for discrete probabilistic modeling. Grad students are also required to implement a naïve Bayes classifier. As usual, please provide your answers in form of a Jupyter Notebook.

Submission. Please submit your assignment as [ML_Assignment3](#) on Brightspace.

Submission deadline. Tuesday, Feb 27, 2:25 pm.

Late submission policy. No late submission will be excepted.

Academic Integrity. Dalhousie academic integrity policy applies to all submissions in this course. You are expected to submit your own work. Please refer to and understand the academic integrity policy, available at <https://www.dal.ca/academicintegrity>

If you have a question: Teaching Assistants (TAs) will be present during the labs to help you with any questions you may have. If you still have questions, feel free to email me at tt@cs.dal.ca.

Questions:

1. **[30 marks]** This Assignment requires you to write a Python program that loads the *Iris* dataset from the first assignment and apply EM with a Gaussian mixture model on the IRIS data. You are not allowed to use any Python library for the EM algorithm itself, but you are of course allowed to use other helper functions. You might compare the results of your program with sklearn models, but the whole exercise is to write the algorithm yourself.
 - 1.1. **[30 marks, 15 marks for Grads]** Write a program to implement EM with a Gaussian mixture model on the *Iris* dataset for $k=3$ and plot the Sepal data points with a color coding based on the obtained clusters. More specifically, you can plot the data points with color where the RGB colour values correspond to the probability estimates of a data point belonging to each class.
Hint: You can use `numpy.linalg.pinv` to find the inverse of a matrix. Also, `numpy.copy` can be used to temporarily save a vector.
 - 1.2. **Graduate students only [15 marks]** Evaluate the prediction quality with different number of assumed classes ($k=2,3,4$). Explain briefly your evaluation method and discuss your finding.
2. **[20 marks, 10 marks for Grads]** This Assignment requires you to write a Python script to calculate some inference of a simplified version of the car repair example from the manuscript. Given are the following probabilities:
The marginal probability that the alternator is broken is $1/1000$ and the marginal probability that the fan belt is broken is $2/100$. The probability that the battery is charging when either the alternator or the fan belt is broken is zero. However, even if both are working there is a $5/1000$ probability that the battery is not charging. When the battery is not charging then there is a 90% chance that the battery is flat, though even if

the battery is charging then there is a 10% chance that the battery is flat. Finally, the car does not start if either the battery is flat, or there is no gas, or the starter is broken. However. Even if these three conditions don't hold there is a 5% chance that the car won't start.

- 2.1. Draw the causal model of this system.
- 2.2. What is the probability that the alternator is broken given that the car won't start?
- 2.3. What is the probability that the fan belt is broken given that the car won't start?
- 2.4. What is the probability that the fan belt is broken given that the car won't start and the alternator is broken?
- 2.5. What is the probability that the alternator and the fan belt is broken given that the car won't start?

Hint: You might use `Lea` methods.

3. **Grads, only [20 marks]** Naïve Bayes:

This Assignment requires you to write a Python program to test a simple binominal version of the Naïve Bayes algorithm on the 20newsgroups dataset. You need to read the data and work with sparse data in python. You should write a Naïve Bayes program on your own (not using library function) to implement the binomial version of the Naïve Bayes rule outlined in the manuscript. Please provide the results in form of a confusion matrix.