Tutorial 8 (10 pt.) Variational Inference on Gaussian Mixture Models using python

Submit Assignment

Due No Due Date **Points** 10 **Submitting** a file upload
File Types pdf, txt, zip, gzip, ipynb, and py
Available Jan 14 at 12am - Mar 13 at 11:59pm about 2 months

This is an advanced tutorial that will give you agood understanding of ELBO and how it is used.
10-20 hours. It is in python and should be straightforward to run. If your calculus is weak do not attempt this.

To possibly understand this one must read up on variational methods. Similar Gaussian mixture models are described in literature. Take a look at this:

Variational Bayesian methods - Wikipedia.pdf

https://en.wikipedia.org/wiki/Variational_Bayesian_methods#A_more_complex_example
(https://en.wikipedia.org/wiki/Variational_Bayesian_methods#A_more_complex_example)

http://bjlkeng.github.io/posts/variational-bayes-and-the-mean-field-approximation/
(http://bjlkeng.github.io/posts/variational-bayes-and-the-mean-field-approximation/)

and or the papers and slides:

bishop-aistats01.pdf

IDAPISlides17_18.pdf

variational-intro.pdf

Here is the tutorial and code:

tutorial.pdf

vi.py

Grading: You must include plots showing the results of running the code showing the resulting clustering. Also to pass you must show you understood both the general theory and the specific GMM example.

You can earn 6 to 10 points depending on how well you do.

Here are my tips from doing the tutorial:

ELBO is often written with a symbol script L(Q) and is also called the (negative) variational free energy

When k is 2 and n is 1, the multinomial distribution is the Bernoulli distribution. When k is 2 and n is bigger than 1, it is the binomial distribution. When k is bigger than 2 and n is 1, it is the categorical distribution.

So the Categorical distribution is the Multinomial Distribution with more than 2 'possible outcomes' and only one trial. So for example throwing a die once has k=6 and n=1.

Also the Dirichlet distribution is the conjugate prior.

So if the prior is Dirichlet and the conditional probability is Categorical then the posterior is Dirichlet again.

Note that the Wikipedia Complex example is like Olgas but more complex.

so

olga -> Wikipedia

alpha -> mu_0
lambda^2 I -> Lambda_(z_i^(-1)
sigma^2 I -> Beta_0 Lambda_i^(-1)
c->z
c_i^T mu -> mu_(z_i)
K-> K
N-> N

Also she assumes the right form for the mean field components without deriving that this is the best.

For assignment 1 work out one term at a time. They turn out to not be so bad if you do it right. For example the first one is essentially the expectation of the log of a normal dist. The only thing that is not the expectation of a constant (wrt mu and c) is the mu term which is a quadratic polynomial in mu. As the q(mu) distribution is normal it is easy to write down the first 3 moments. (ie 1, m_k, s_k^2+m_k^2)

Do not forget the normalization term. It might matter for some of these.

Notice that P(c_i) = 1/K no matter what c_i is. So that expectation is easy.

Third one is the most exciting as it has both mu and c in it. Have fun.

For assignment 2 it is nice that one can just do the partial derivative of the ELBO from assignment 1 wrt phi_{ik}, set it to zero and solve for phi_{ik}

The result is then the same when you notice
E[mu_k]=m_k and E[mu_k^22]=m_k^2+s_k^2

Wow this all seems consistent.

Assignment 3 Complete the square means add and subtract constant terms and regroup so that it looks like A(mu_k-B_k)^2 +C for some A, B_k and C. Sort of obvious what A is since it has to multiply mu_k^2 so just match those quadratic terms. The C is not important and the B is not hard either.