

# Accelerating your Python Code

## With PyCuda and GMMs

Varun Nayyar

27/07/18

# Outline

Me = Math Major + Script Kiddy (Manage Expectations please)

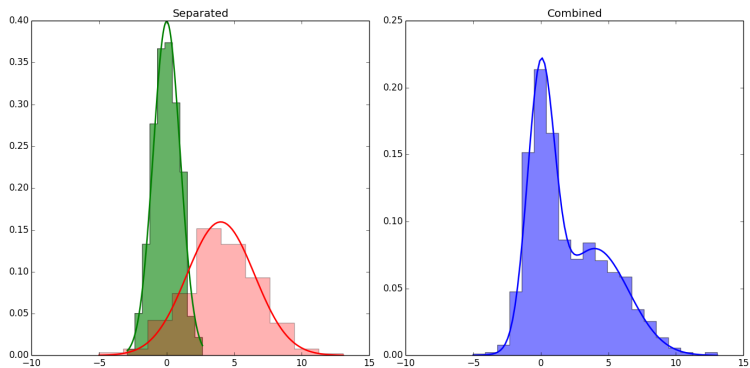
## What to Expect

- Some Math
- Iterative Process I went through
- Thinking with CUDA
- How to use PyCUDA to avoid complicated work

## What not to Expect

- How PyCUDA does it's magic
- Advanced CUDA (or even Intermediate)

# Gaussian Mixture Models $d=1$ , $K=2$



# Gaussian Mixture Models (GMMs)

## Mathematical Notation

Gaussian Mixture Models are a class of models that assume the data points are generated by a finite number of gaussian (or normal) distributions

$$f(\mathbf{x}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{k=1}^K \pi_k \mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k)$$
$$l(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{x}) = \sum_{i=1}^N \ln \left( \sum_{k=1}^K \pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k) \right)$$

with  $K$  mixtures and the weights  $\pi_k$  giving the prior probability of mixture component  $k$  (and sum to 1).

# Doing it Live

This can't



## Need for speed

- GMM likelihood formula doesn't decompose into a mathematically simpler form
- Number of flops are of the order of  $O(NKd)$ , in my case,  $N = 100,000$ ,  $K = 8$ ,  $d = 13$ .
- I needed to evaluate the likelihood 100,000 times for a fixed dataset while the parameters were varied.