Kernel Density Estimation with Mixture of Gaussians

Dhananjay Singh

June 2019

1 Introduction

This work implements a model derived from kernel density estimation, formulated in terms of mixture of Gaussians.
The model as defined in the problem statement:

$$\log p(x) = \log \sum_{i=1}^{k} \exp\{\sum_{j=1}^{d} [-\frac{(x_j - \mu_{i,j})^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)] - \log k\}$$

This can be simplified as

$$\log p(x) = -\log k + -\frac{d}{2}\log(2\pi\sigma^2) + \log \sum_{i=1}^{k} \exp\{\sum_{j=1}^{d} [-\frac{(x_j - \mu_{i,j})^2}{2\sigma^2}]\}$$

To improve the efficiency and execution time, the following strategies were adopted:

1. Loops have been reduced by using vectorization.
2. To fully exploit processing capabilities, data was divided and executed in parallel among multiple-processors.
3. Image precision was reduced from float64 to float32, cutting memory and execution time. This approach was verified to have negligible differences in output.

2 Dataset

This section contains random images obtained after processing dataset.

2.1 MNIST

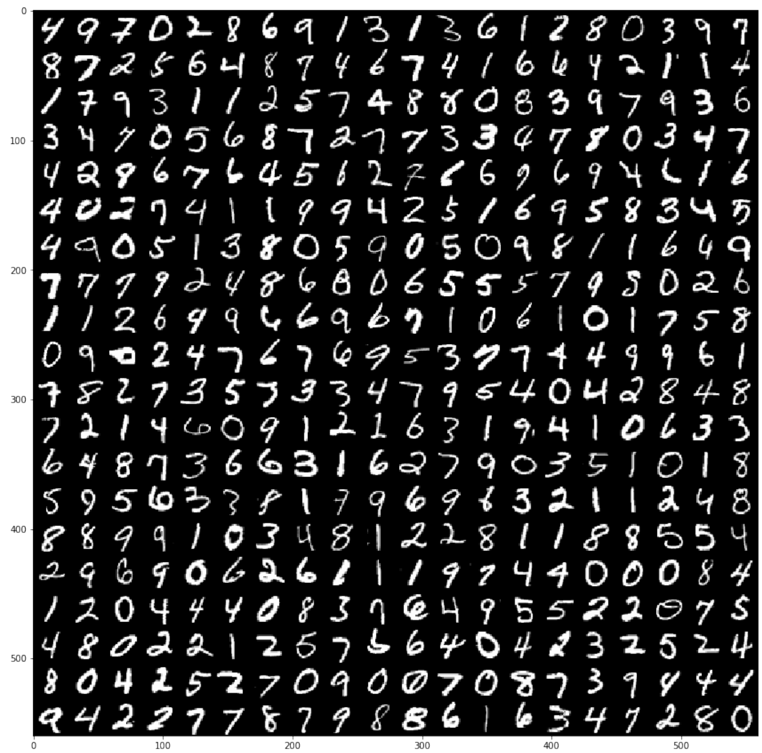

Figure 1: 400 random samples from the MNIST training split.

2.2 CIFAR

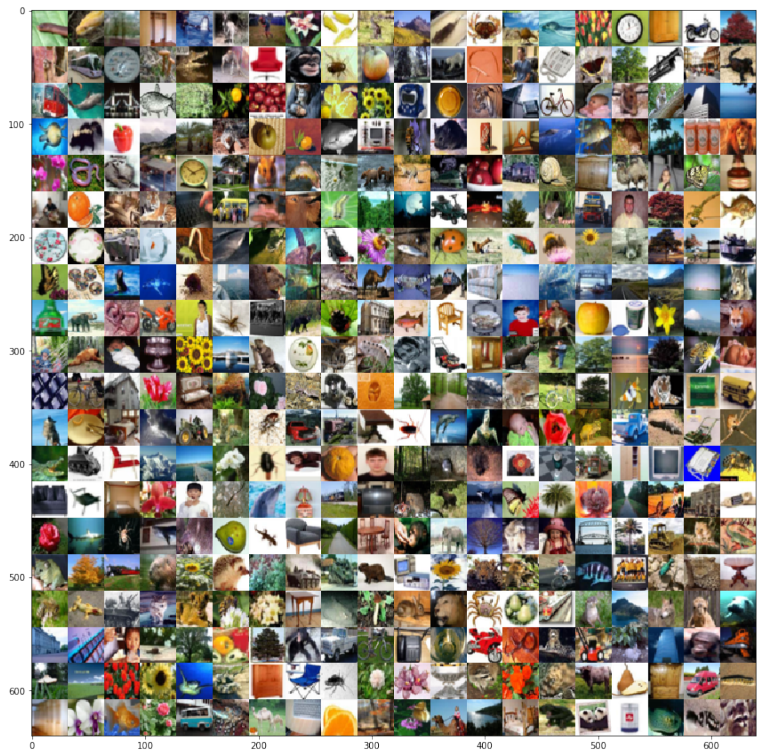

Figure 2: 400 random samples from the CIFAR 100 training split.

3 Experimental Results

3.1 Finding Optimal Hyperparameter

We run grid search on the validation split of each dataset and establish the optimum value of σ.

3.1.1 MNIST

Table 1 captures the results of the grid search on MNIST. We find that $\sigma = 0.2$ houses the optimum value.

Table 1: Gridsearch on MNIST

σ	log likelihood
0.05	-3178.95
0.08	-623.68

Table 1: Gridsearch on MNIST

σ	log likelihood
0.1	-123.91
0.2	**232.29**
0.5	-234.05
1.0	-740.94
1.5	-1051.15
2.0	-1272.98

3.1.2 CIFAR100

Table 2 captures the results of the grid search on CIFAR100. We find that $\sigma = 0.2$ houses the optimum value.

Table 2: Gridsearch on CIFAR100

σ	log likelihood
0.05	-13773.63
0.08	-2942.02
0.1	-794.73
0.2	**852.99**
0.5	-904.21
1.0	-2882.26
1.5	-4099.42
2.0	-4972.69

3.2 Verifying Results

Dataset	Log Likelihood	Time Taken
MNIST	234.28	1.67 min
CIFAR100	843.76	7.41 min

3.3 Analysis

3.3.1 Increasing σ

We see that initially increasing the σ increases the likelihood but only till a point after which the likelihood dips. In a sense this is due to the density shape going from under-smooth to over smooth. A lower value of sigma leads to the closer μ (closer pixel value) to be weighted higher. For a value of σ in between the two extremes, the density estimated is closest to true distribution, leading to a higher likelihood.

3.3.2 Model Compatibility

Pros of the Model:

- Doesn't require a prior assumption of the distribution of the data. e.g. For Image tasks where background is useful, we can estimate the background distribution.
- For some kind of images, can be useful to get a sense of the distribution. (More in cons)

Cons of the Model:

- Images can be complex. This model inherently memorizes data points and the estimated density may not generalize to unseen points. e.g.Same object different colors than those present in data. The solution would be to gather more data points, to be more representative.
- As number of features go up, the memory/compute requirements increase drastically. For high dimensional data this makes the model expensive. But then again, we are living in the age of deep learning scale of computation :)