Kernel Density Estimation using a Gaussian Kernel

By: Valeriu Tocitu

This code uses Kernel Density Estimation (KDE) to classify images. The Kernel used is a Gaussian Kernel whose log-probability is given by the following equation

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

Where

- k = the size of the training data
- d = the size of the validation data
- x_j = the training data
- $\mu_{i,j}$ = the validation data.

Equation 1 simplifies to

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

Using the log-sum-exp trick,

$$L = -\frac{d}{2}\log(2\pi\sigma^2) - \log(k) + a + \log \sum_{t=1}^{k} \exp\{\sum_{j=1}^{d}[\frac{(x_j - \mu_{i,j})^2}{2\sigma^2} - a]\}\text{, (3)}$$

Where

$$a = max_t\{\sum_{j=1}^{d}[\frac{(x_j - \mu_{i,j})^2}{2\sigma^2}]\} \text{ (4)}$$

We demonstrate this on the MNIST and CIFAR100 datasets. An example of each dataset is shown in the images below.

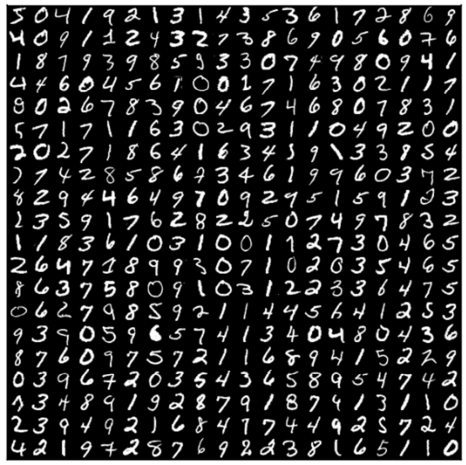

Figure 1: MNIST Sample

Figure 2: CIFAR100 Sample

In both cases, the first 10K samples of the training dataset remains as such, while the second 10K samples become the validation dataset. The rest of the samples are discarded.

σ	L_{MNIST}	$L_{CIFAR100}$
0.05	-3102.1340289273339	-13082.488664599698
0.08	-593.67488965287055	-511.03227516674405
0.1	-104.70815449386602	-327.05956793810868
0.2	237.09484316652816	918.92899042919112
0.5	233.27711137696383	-693.63103532460252
1	-740.72318186930841	-2822.9791740047544
1.5	-1051.0042893879461	-4068.5679861130357
2	-1272.8624506765473	-4952.3273126849072

Table 1: Log-probability vs. Value of Sigma for both datasets.

We conduct a grid search to find the best possible sigma. Table 1 above shows the log-probability for each dataset associated with each sigma value. From Table 1, we see that the optimal σ = 0.2. That is the σ associated with the minimum positive L. Replacing the validation data with the test data, Table 2 shows the log probability of each dataset.

Dataset	L_{max}
MNIST	861.406205894
CIFAR100	228.198219724

Table 2: Optimal Log-probability vs. Value of Sigma for both datasets.

Conclusion

This data demonstrates that a KDE with a Gaussian Kernel is useful in classifying images. The MNIST dataset has a higher log-probability and a faster runtime. This is to be expected since the dataset consists of grayscale images. In addition, there are only 10 possible categories. By comparison, the CIFAR100 dataset consists of color images classified into 100 categories.