Langevin Dynamics in K-SVD

Anonymous Authors[1]

1. Introduction

Sparse Coding attempts at learning basis functions for an ensemble of images arising from the same process and inferring sparse representations using the dictionary components (atoms). It assumes a generative model with additive Gaussian noise, an overcomplete dictionary, and a sparse (Laplacian) prior on the coefficient vector. This model fails to work for low light (LL) images because of its low contrast and signal-to-noise ratio, as well as the inherent discrete nature of limited photon counts [2][3].

The goal is to recover well-illuminated ground truth from Low Light images. In the sampling regime, however, intractability of the L0-type optimization becomes tractable. Fang and Olshausen adapted Langevin Sampling to Gaussian Sparse Coding. We intend on applying it to K-SVD in the BARS and MNIST dataset to yield more accurate solutions. We will also be able to sample the posterior distribution over latent variables.

We demonstrate this idea for a sparse coding model by deriving a continuous-time equation for inferring its latent variables via Langevin dynamics. The model parameters are learned by evolving alongside according to another continuoustime equation, thus bypassing the need for digital accumulators or a global clock. We also introduce a novel method for imposing a sparse prior that encourages latent variables to be exactly zero (so-called 'L0 sparsity'), in contrast to the relaxed L1 cost normally utilized in sparse coding models. This gives rise to an efficient procedure for sampling from the posterior, bypassing the problems normally associated with gradient-based inference in this regime. Simulations of the proposed dynamical system on both synthetic and natural image datasets demonstrate that the model is capable of probabilistically correct inference, enabling learning of the dictionary as well as parameters of the prior.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

1.1. Formulation

Gaussian Sparse Coding

The problem of encoding sensory information efficiently is relevant both to the design of practical vision systems and to advancing our understanding of how biological nervous systems process information. Within the image processing community, much work has been done on image codes that utilize a linear basis function expansion, and considerable effort has gone into choosing sets of basis functions that satisfy certain mathematical desiderata or that have desirable properties, such as ease of computability.

We assume a generative model for sparse coding:

$$x = As + \varepsilon \text{ s.t } \varepsilon \sim N(0, \sigma^2 I)$$

A represents an overcomplete dictionary with each column being called atoms. s must be as sparse as possible. This can be achieved by setting some prior peaked at zero for $P(s)$ such as the Laplacian distribution which leads to the formulation of the optimization problem:

$$\min |As - x|^2 + \lambda |s|_1$$

The goal is to find a dictionary whose columns are basis functions, A. And s whose entries are basis coefficients. In an ideal world, we would want to use the L0-norm for the optimization problem but this is computationally intractable in most cases.

Langevin Dynamics

Langevin Sampling enables us to use L0-like priors tractably. We are specifically interested in using this technique to recover the dictionary of a set of images. There are tons of papers on sparse coding to learn dictionaries of BARS/MNIST dataset. This will provide us a plethora of techniques to compare against when assessing the effectiveness of Langevin Sampling.

Below, we have the equation that governs Langevin Dynamics:

$$\frac{dx}{dt} = -\nabla E(x) + \eta(t)$$

It is essentially gradient descent with additive Gaussian Noise that is uncorrelated in continous time as illustrated here:

$$\langle \eta(t), \eta(t') \rangle = 2\delta(t - t')$$

It is proven that the distribution eventually converges to:

$$p^{\infty}(x) \propto exp(-E(x))$$

We can also define second-order Langevin Dynamics which includes a momentum based parameter that improves smoothness of convergence.

$$m\frac{d^2x}{dt^2} + \frac{dx}{dt} = -\nabla E(x) + \eta(t)$$

Langevin Dynamics also takes into account time-varying energy functions and still works. We leverage this by allowing latent variables to be the coefficient vector α that we are trying to learn alongside dictionary D.

$$\begin{aligned}&\frac{dx}{dt} = -\nabla E(x, u(t)) + \eta(t)\\&\langle e^{-E(x,U)} \rangle_{U \sim p_U}\\&e^{-\langle E(x,U) \rangle_{U \sim p_U}}\end{aligned}$$

K-SVD

K-SVD is a dictionary learning algorithm to find sparse representations of data. The formulation is as follows:

$$\min_{X,D} ||DX - Y||_F^2 \text{ subject to } ||x_i||_0 \leq T$$

The problem is solved using alternating minimization:

$$\begin{aligned}&X^{(k+1)} = \min_X ||D^{(k)}X - Y||_F^2 \text{ subject to } ||x_i||_0 \leq T\\&D^{(k+1)} = \min_D ||DX^{(k+1)} - Y||_F^2\end{aligned}$$

Which alternates between updating the sparse representation X and the dictionary D.

2. Methods

Langevin Dynamics in Gaussian Sparse Coding

We normally use iterative linear algebra based optimization techniques like gradient descent to solve the L1 regularized optimization problem. But this is only a convex envelope and is not exact so we forfeit accuracy for efficiency. In the probabilistic setting an L1 cost corresponds to a Laplacian prior, which only weakly captures the notion of sparsity. The problem of the L0 norm being intractable disappears in the sampling regime. Moreover, sampling can be faster than optimization for high-dimensional non-convex problems.

Sparse coding is a simple yet efficient algorithm for learning structure in data by finding a 'dictionary' to describe patterns contained in the data. While it is formulated as a probabilistic latent-variable model, it is often approximated in practice by finding point estimates for the latent variables rather than sampling from their posterior distribution. As a result, it is difficult to make rigorous claims about the relation between the learned dictionary and the statistics of the data, and it is problematic to adapt other parameters of the model such as the degree of sparsity or overcompleteness of the dictionary.

K-SVD update with Langevin Dynamics

Algorithm 1 Langevin K-SVD

Input: data x_i, size m
for $k \leftarrow 1$ to N_A **do**
  $X \leftarrow$ SAMPLEBATCH()
  for $n \leftarrow 1$ **to** N_s **do**
    $s \leftarrow s - \eta_s \cdot \nabla_S E(A, s, x)$
  end for
  $s^* \leftarrow s$
end for

3. Results

3.1. Bars Dictionary

We call the combine the K-SVD algorithm with Langevin Sampling KLS. We evaluate KLS on the bars dataset. The bars dataset is shown in Figure (1).

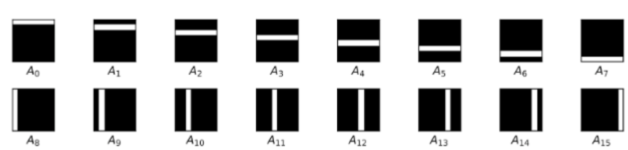

Figure 1. Bars Dictionary

The synthetic noisy data Y is defined as

$$Y = DX$$

Where the columns of $Y = [\vec{y_1} \cdots \vec{y_n}] \in R^{d \times n}$ are given by

$$\vec{y_i} = D_{true}\vec{x_i} + \mathcal{N}(\vec{0}, \sigma^2 I).$$

It remains our goal to $D \in R^{d \times k}$ and $X \in R^{k \times n}$ from Y. For the bars dictionary, we would like each column of the learned dictionary to have a bar like form (shown in Figure (1)). Using KLS, we obtain the results shown in Figure (4).

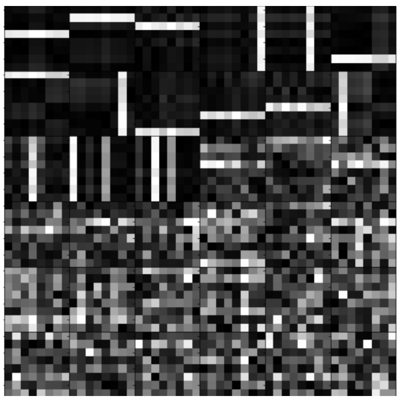

Figure 2. KLS-learned Bars Dictionary

We can see that KLS successful learns the characteristic bars. This toy example verifies that KLS indeed works as intended.

3.2. MNIST Dictionary

In order to further evaluate the performance of KLS, we use the MNIST dataset to learn a dictionary for handwritten digits. The entire 28×28 pixel digits are used as data vectors of length $d = 784$. We train on $n = 5000$ digits in order to learn $k = 100$ atoms in our dictionary D. Figure (**??**) shows the resulting learned dictionary. As we can see, the unnecessary dictionary atoms look like noise. The reason for this is that the dictionary is intially set to randomized values, and is iteratively updated one atoms at a time. Thus, any unnecessary atoms will not be effectively updated.

3.3. KLS Low Light

Finally, we evaluate our KLS method on low-light images. Low light images will contain much more Poisson noise. This is because in the low light regime, photon counts are much lower and per pixel, it is small enough to be considered discrete. A lot of the current techniques involve approximating the Poisson Noise as Gaussian Noise through Anscombe transform and then conducting regular sparse coding techniques on that. This is not viable for Poisson means less than 5 which is the case for Low Light regimes.

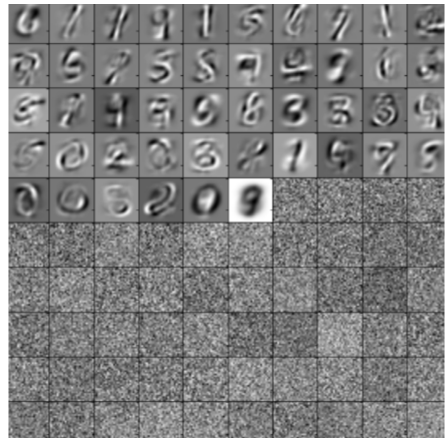

Figure 3. KLS-learned MNIST Dictionary

Using the open dataset of low light images [1]. To really observe the results of the denoising algorithm, we zoom in on a 200×200 pixel region of a natural low light image. The results are shown in in Figure (**??**).

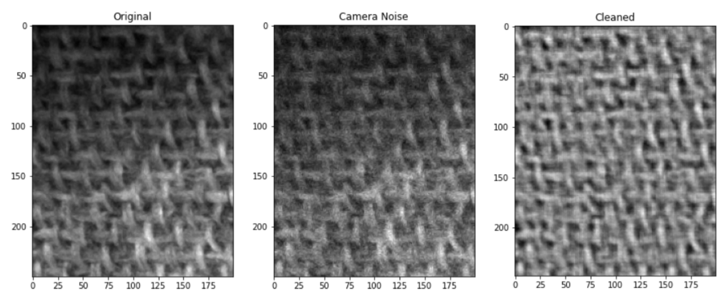

Figure 4. KLS-learned MNIST Dictionary

We notice two main things. First, the noise seems to have been attenuated. Additionally, the image's vertical darkness variation has not been preserved. Our method seems to have removed slow variations in the image while removing the noise. We believe this may be fix with more parameter tuning (patch size, stride length, number of iterations, number of atoms, etc). However, the results highlight the capability for this algorithm to remove noise from low light images.

3.4. Evaluation

Unfortunately, we were not able to achieve the desired accuracy for the MNIST dataset. BARS was reasonably good. This is likely because the dictionary is far more rudimentary, and we need to figure out ways to improve feature extraction capabilities of our model. Some possible extensions of the project include applications to medical imaging (CT for example) where low light scenarios arise often.

Link to code: https://github.com/danielabrahamgit/sparse_coding

References

[1] Josue Anaya and Adrian Barbu. "RENOIR - A Dataset for Real Low-Light Noise Image Reduction". In: *Journal of Visual Communication and Image Representation* 51 (2018), pp. 144–154.

[2] Joseph Salmon et al. *Poisson noise reduction with non-local PCA*. 2014. arXiv: 1206.0338 [cs.CV].

[3] Lei Liu Xiang-Yu Kong and Yun-Sheng Qian. *Sparsity Based PoLow-Light Image Enhancement via Poisson Noise Aware Retinex Model*. Tech. rep. IEEE, 2021.