CSE516/ECE559: Theories of Deep Learning
Problem Sheet 2

Dr. Vinayak Abrol

September 11, 2024

Question 1 should be submitted for marking. The remaining questions entail computational experiments can be attempted in programming language of your choice with the link to your code and results.

Coding setup:

1. Look at SPAMS toolbox for a collection of Dictionary Learning/Sparse Coding algorithms. `http://spams-devel.gforge.inria.fr/`, but you are free to use any online resource.

---

1. Prove the Dictionary Learning minimization problem can be solved by an alternate maximization objective

$$\min_{D,A} \|X - DA\|_F^2 = \|X\|_F^2 - \max_{D} \sum_{n} \max_{|I| \leq S} \|D_I D_I^{\dagger} x_n\|_2^2$$

   where S denotes the cardinality of column vectors a_n of matrix A, $\|a_n\| = \|D_I^{\dagger} x_n\|$ as a result of employing iterative thresholding type projection algorithm for sparse coding with $\dagger$ denoting the pseudo-inverse.

2. Go to playground.tensorflow.org and choose the spiral dataset. Now, construct a **minimum width and depth** Relu network which can approximately solve this classification problem. Include, the snapshot of your screen after your training is successful and you feel that this is the best you can do.

3. **Dictionary Learning**

   (a) Using Fashion-MNIST dataset, learn a KSVD dictionary, ICA basis and PCA basis. Choose hyperparameters of your liking for KSVD and compare the results.

   (b) Predictive sparse decomposition and autoencoders: Computing sparse code for a signal (assuming ℓ_0 or ℓ_1-norm based constraint) is slow. Let's try to train a simple predictor to compute an approximation to the sparse code as:

$$\|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \alpha\|\mathbf{x} - \mathbf{F}(\mathbf{y}, \mathbf{W})\|_2^2$$

   .

   The hope is that, after training, the value of $\mathbf{x}$ can be inferred directly through $\mathbf{x} = \mathbf{F}(\mathbf{y}, \mathbf{W})$, instead of running a tedious sparse coding optimization procedure. Train this autoencoder using the following function $\mathbf{F}(\mathbf{y}, \mathbf{W}) = \log(1 + \exp(\mathbf{W}\mathbf{y}))$ on Fashion-MNIST dataset. The function $\log(1 + \cdot)$ is applied to each component of its vector argument and can be seen as a smooth version of ReLU. You have to implemented an alternating procedure in training $\mathbf{D}$ and $\mathbf{W}$, but it is also okay to use the tools available in the SPAMS package.

   Now comment and try out the encoder $\mathbf{F}(\mathbf{y}, \mathbf{W}) = \max(0, \mathbf{W}\mathbf{y})$.