Student ID : 20194293

Name : Go, Kyeong Ryeol

[AI 502] k-Sparse Autoencoders

1. Paper Summary

In classification task, it has been shown that encouraging sparsity improves performance. In previous works, it is accomplished by sparse coding approach or sparsity penalties in neural network, however, these suffer from the computational complexity and other requirements such as combinations of activation functions and sampling steps although the sparse representations of each input is not guaranteed sometimes. Therefore, the author devised a novel model, which is named as "k-Sparse Autoencoder", that can be efficiently learnt and used for sparse coding where the sparsity is the only regularizer and the only non-linearity.

k-Sparse Autoencoder is a simple variant of basic autoencoder with linear activation functions and tied weights. In the feedforward phase, after computing the hidden code $\mathrm{z} = W^T x + b$, the k largest hidden units are selected and set the others to zero. This can be done by sorting the activities or by using ReLU hidden units with thresholds that are adaptively adjusted until the k largest activities are identified. The selection step acts as a regularizer that prevents the use of an overly large number of hidden units when reconstructing the input. Empirically, at test time, slightly better performance is obtained by using the αk largest hidden units where α is tuned via validation set.

As learning, the author proposed Iterative Thresholding with Inversion (ITI) to find the sparsest solution of $\mathrm{x} = \mathrm{Wz}$. This consists of 2 steps which are support estimation step $(\Gamma = \mathrm{supp_k} * z^n + W^T(x - Wz^n))$ and inversion step $\left(\mathrm{z}_\Gamma^{n+1} = \left(W_\Gamma^T W_\Gamma\right)^{-1} W_\Gamma^T \mathrm{x} \;\;\&\;\; \mathrm{z}_{(\Gamma)^c}^{n+1} = 0\right)$. Here, after estimating the support set of z as Γ, then W is restricted to the indices included in Γ and form W_Γ. Then, using the pseudo-inverse of W_Γ, the value of z that minimize $||x - W_\Gamma z_\Gamma||_2^2$. Finally, the support estimation is refined and this process is repeated until convergence.

To qualitatively analyze the validity of k-Sparse Autoencoder, the filters of the encoder are plotted varying the sparsity level. It is shown that for large values of k, the algorithm tends to learn very local features whereas the global features were learnt with small k. Nevertheless, since forcing too much sparsity results in too local features, level of sparsity should be elaborately selected depending on the task. Furthermore, in both supervised and unsupervised learning task, k-Sparse Autoencoder outperforms the other models in terms of error rate. Moreover, when plotting the log-histogram of hidden unit activities, the activation values were far more diverse in case of the proposed model while the hidden unit activation values were in some sense concentrated in low values for ReLU Autoencoder and Dropout Autoencoder.

2. Discussion

Here I would like to offer 2 discussion points. To begin with, can L1 regularization show the identical performance? Since LASSO is traditionally well-known method for feature selection. Therefore, there is something in common with ITI in the sense that the weight matrix is adapted to the indices that is included in the support set. I think it's possible if the coefficient of L1 regularizer in the loss is properly tuned. Next, what if non-linear activation function is incorporated to the model? In my opinion, tanh or sigmoid function may be utilized as these does not change the order of the hidden unit activities so that the selected indices do not change. However, it needs an assumption that these activation does not occur the gradient vanishing problem. In that sense, batch normalization may be helpful.