Parent Item: Learning Convolutional Nonlinear Features for K Nearest Neighbor Image Classification

Summary

Intro

- They aim to find a way to learn better feature representations that are tailored specifically for KNNs
- Use NCA as loss function and CNN to learn features

Related Work

- This paper references a lot of good possible metric learning algorithms (NCA, Mahalanobis distance, RCA, ITML, RBM, LMNN, etc)
- Salakhutdinov paper references extends NCA to be nonlinear, possible other paper to read through for understanding
- Their difference between that ^ and their work is that their model learns discriminatively trained convolutional nonlinear features and model natural images better
- They also say their model can handle more complex image data and larger datasets

Proposed Method

- Want to build a CNN with the training objective to optimize KNN classification error (using NCA as the loss)
- This loss function is differentiable and continuous so it could be useful as our loss function for our model as well since it is also tailored for KNN
- They use Stochastic Gradient Descent as their learning algorithm
- For their hyperparams: They need to keep shuffling the batches to avoid overfitting (standard), they monitor their training of the model to change the learning rate (mention .01 or .001 being good), apply early stopping, and use momentum learning.

Experiments

- Use MNIST and CIFAR-10 datasets to test classification for the KNN
- Again, they use fully connected layers at the end of the convolutions which is how this differs from our work. It seems that they only use the CNN for feature learning and input those features into the KNN separately so the KNN is not replacing fully connected layers in the end-to-end process
- Use tanh activation for each layer
- Also trained another CNN with the same architecture using cross entropy loss instead for comparison of the features' performance
- They conclude that this method is an effective way to improve KNN classification tasks and also varify with visualizations that the CNN learned useful features and is able to separate out the classes into more distinct neighborhoods than other standard methods

Conclusion

- Able to learn better features and NCA as loss helps more specifically to KNN classification
- Some problems: NCA is non-convex so it suffers from local-optima and learned features are real valued vectors thus not efficient in larger scale applications