Nathan Subrahmanian
COSI-159a
Due: 3/15/24

What are the contributions of this paper?

There are 3 major contributions that the SphereFace paper makes. First, they propose using the A-softmax (angular softmax) loss for CNNs to learn discriminative face features with a novel geometric interpretation. These features lie on a hypersphere manifold, matching previous hypotheses that faces themselves will lie on a manifold of the hypersphere. The second contribution is that the angular margin can be adjusted via a parameter m*, for which the authors derived lower bounds for such that the loss approximates the learning task that minimal distance between classes is larger than the greatest intra-class distance. Their third important contribution is to show that using angular margins are effective for the face recognition task by achieving high performance on multiple datasets across several benchmarks.*

Illustrate the three properties of the proposed A-Softmax.

- *1) The features learned through Softmax loss are angularly distributed instead of linearly (in normal softmax) and with adjustable difficulty*
    - *As m increases, the angular margin becomes larger, the constrained region of the manifold shrinks, and the learning task becomes more difficult*
- *2, 3) There is a minimal optimal m, for which the authors have derived some lower bounds*
    - *2) When doing binary classification, the lower bound for m >= 2 + sqrt(3)*
    - *3) When doing multi-class classification, the lower bound is m >= 3*
- *Introduces a geometric constraint such that it (the A-softmax) is sensitive to the angular relationship between feature vectors and the centroids for a class rather than just Euclidean distance*
    - *This is better for tasks like face recognition where there is a lot of intra-class variety in data*

What is the evaluative metric used for the LFW dataset? How is it calculated?

The evaluative metric used in training is the cosine distance between features in different training examples from the same class (person). We can look at the angular margins between positive and negative pairs and their distributions, which tells us that, as the hyperparameter m is increased, so is system accuracy. The downstream evaluation metric is just accuracy (%) for both LFW and YTF. Accuracy is just calculated by # correct predictions / # test examples.