Task 1

1. The contributions of the paper are as follows:
    a. The introduction of the Angular Softmax (A-Softmax) loss for Convolutional Neural Networks (CNNs) to learn discriminative face features with clear and novel geometric interpretation. The learned features discriminatively span on a hypersphere manifold, which intrinsically matches the prior that faces also lie on a manifold.
    b. Derivation of lower bounds for the parameter m such that A-Softmax loss can approximate the learning task where the minimal inter-class distance is larger than the maximal intra-class distance.
    c. Demonstration of the effectiveness of angular margin in face recognition, with the trained model, SphereFace, achieving competitive results on several benchmarks including Labeled Face in the Wild (LFW), YouTube Faces (YTF), and MegaFace Challenge 1.
2. The three properties of the proposed A-Softmax are:
    a. A-Softmax loss defines a large angular margin learning task with adjustable difficulty. With a larger value of m, the angular margin becomes larger, and consequently, the constrained region on the manifold becomes smaller, making the corresponding learning task more difficult.
    b. In the binary-class case, there is a lower bound for the minimal m (m_{min}) such that the maximal intra-class angular distance is constrained to be smaller than the minimal inter-class angular distance.
    c. Under the assumption that the weights W_i are uniformly spaced in Euclidean space, there is a lower bound of $m_{min} \geq 3$ in the multi-class case.
3. The evaluative metric used for the LFW dataset is the accuracy percentage. The methodology for evaluating face recognition performance on the LFW dataset involves computing the cosine distance between feature vectors of paired face images and classifying them as either the same person or different people. The standard approach for this dataset includes using the output features from the FC1 layer, applying cosine similarity for scoring, and evaluating performance based on True Positive Rate (TPR) and False Positive Rate (FPR) at various thresholds. The overall performance is then reported as the mean accuracy over multiple testing splits or folds.

Task 2

Task 2 was completed by evaluating and incorporating the six grading criteria - Data loading and augmentation, Neural network design, Loss function, Training, Testing and results, and Code format - based on the methodologies outlined in the paper "Sphereface: Deep hypersphere embedding for face recognition." Below are the specifics of how the paper influenced the implementation of each grading criterion:

1. Data Loading and Augmentation:
The paper advocates horizontal flipping as the sole data augmentation technique. However, to enhance the model's generalization ability, additional data augmentation methods such as resizing, random rotation, random cropping, color jitter, and normalization were adopted.

2. Design of Neural Networks:
According to the paper, the optimal architecture is a CNN, varying in depth from 4 layers to 64 layers. The assignment required the construction of a 4-layer CNN. Based on the paper's guidance, the network was configured with convolutional layers having [3×3 64], [3×3 128], [3×3 256], [3×3 512] filters, each with a stride of 2, and a fully connected layer with 512 units.

3. Loss Function:
The adaptation of the loss function was in accordance with the paper's detailed explanation of the A-Softmax loss, including the recommendation to set the margin parameter, m, to 4, enhancing the feature distribution effectiveness for open-set face recognition.

4. Training:
The paper recommends specific settings for the learning rate schedule and batch sizes. Specifically: a batch size of 128 spread across four GPUs; starting learning rate at 0.1, with reductions by a factor of 10 at the 16K and 24K iteration marks, concluding at 28K iterations.

5. Testing and results:
The evaluation methods recommended by the paper involve using cosine distance between features to score face identification and verification during testing. This includes employing a nearest neighbor classifier and thresholding techniques. The paper's validation of these methods across various standard datasets, including LFW, YTF, and MegaFace Challenge, supports their effectiveness and appropriateness for assessing the results.

The implementation of the "Sphereface: Deep hypersphere embedding for face recognition" model was evaluated in two distinct scenarios: first, against the accuracy of the TA's provided model; and second, by incorporating additional data augmentation techniques into the Sphereface model. In the comparison without data augmentation, the Sphereface model achieved a peak accuracy of 70.00% on the LFW dataset, significantly surpassing the TA's model, which achieved a 65.80% accuracy. This superior performance underscores the effectiveness of the Sphereface architecture and its unique angular penalty loss in neural network training. Furthermore, enhancing the Sphereface model with data augmentation techniques, including resizing, random rotation, random cropping, color jitter, and normalization, resulted in an improved accuracy of 74.20%. This indicates that the Sphereface model's performance can be further enhanced through strategic data augmentation.