NAME=KHUSHI JAIN, 22075042
KORADA RESHMA, 22075044
GROUP NUMBER-19
AM SOFTMAX

Additive Margin Softmax (AM-Softmax) loss is a loss function that adds a margin to the target logit (raw/unnormalized output), which can increase the angular margin between classes. It is used for multi-classification in neural networks. It is used for deep learning tasks, particularly in the field of face verification. Imagine if you're trying to teach a computer to recognize faces. You show it lots of pictures, and for each one, you tell it who's in the picture. This is what we call training the computer. Now, the way the computer learns to recognize the faces of individuals by drawing out some patterns, by using something called a loss function. The face verification task can be viewed as a metric learning problem, so learning large-margin face features whose intra-class variation is small and inter-class difference is large is of great importance in order to achieve good performance and this is where AM-Softmax loss function plays its role.

Difference between AM Softmax and Softmax:

Softmax and AM-Softmax (Angular Margin Softmax) are both loss functions used in the context of deep learning for classification tasks. They have distinct purposes and characteristics:

Softmax Loss:

- **Purpose:**The softmax loss is commonly used in multi-class classification problems, where an input is assigned to one of several mutually exclusive classes.

- **Function**: The softmax function takes the raw output scores (logits) from a neural network and converts them into probabilities. These probabilities represent the likelihood of the input belonging to each class.

- **Characteristics**: It enforces that the sum of the predicted probabilities for all classes is equal to 1, making it suitable for problems where each input belongs to exactly one class.

- **Use Cases:** Commonly used in tasks like image classification, natural language processing (NLP) text classification, and more.

AM-Softmax (Angular Margin Softmax) Loss:

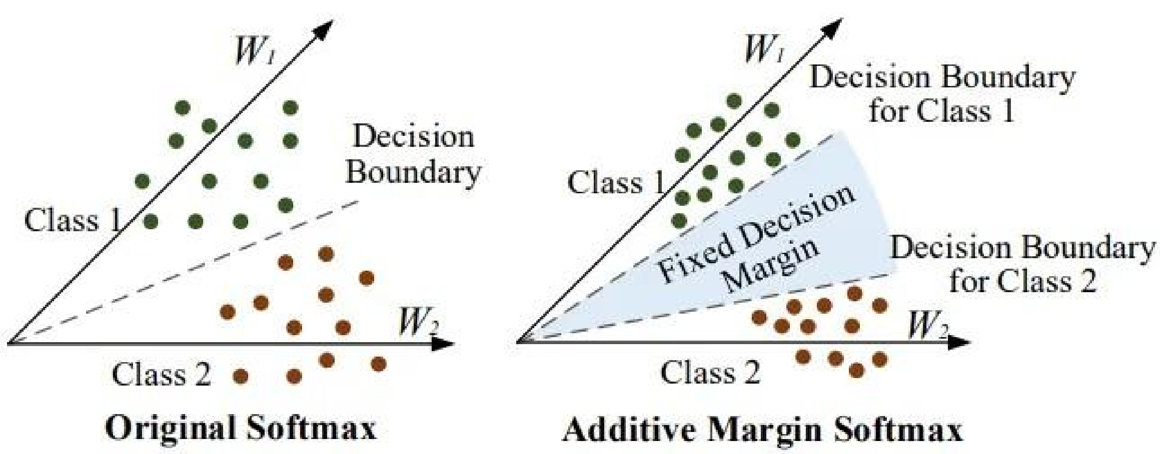
- **Purpose:** AM-Softmax is primarily used in face recognition and similar tasks, where you want to enhance the discrimination between classes. It's often used to learn feature representations that are more suitable for verification and identification tasks.

- **Function**: It introduces an angular margin parameter to the softmax loss. This margin controls the angular separation between feature vectors corresponding to different classes. The goal is to push the feature representations of different classes further apart in angular space.

- **Characteristics**: AM-Softmax encourages the model to produce feature representations that have larger angular separations between different classes while maintaining probabilistic characteristics similar to the softmax loss.

- **Use Cases:** Commonly used in face recognition to learn embeddings for images where each class corresponds to a different person.

In summary, the key difference is that softmax is used for standard multi-class classification tasks, ensuring that each input belongs to one and only one class, while AM-Softmax is used in situations where you want to create distinct feature representations and enhance discrimination between classes, often seen in face recognition and similar problems. AM-Softmax introduces an angular margin to encourage greater separation between classes in feature space.



EXAMPLE:

Consider a face recognition system that needs to distinguish between different individuals. The system is trained on a dataset of face images, where each image is labeled with the identity of the person. The goal is to learn a model that can correctly classify new face images. In this scenario, the softmax loss function would focus on maximizing the difference between the classes (different individuals), but it might not effectively minimize the variation within each class (different images of the same individual). This could lead to a model that performs well on the training data but struggles to generalize to new images of the same individuals. On the other hand, the AM-Softmax loss function introduces an additive margin into the softmax loss, which encourages the model to not only maximize the inter-class differences but also minimize the intra-class variations.

This results in a model that can more effectively recognize different images of the same individual, leading to improved performance on face verification tasks. Let’s say you have a bunch of pictures of your friends and you want to sort them by person. If you use the original softmax function, it might mix up different pictures of the same person. But if you use the AM-Softmax function, it will do a better job of grouping together pictures of the same person, even if they’re in different lighting, angles, or expressions. The AM-Softmax loss function adds a little twist to the softmax function. It introduces a margin, which is a fancy way of saying it adds a buffer or a gap between the classes. This helps to make sure that the pictures of the same person are grouped closer together and the pictures of different people are further apart.

MATHEMATICAL FORMULATION:

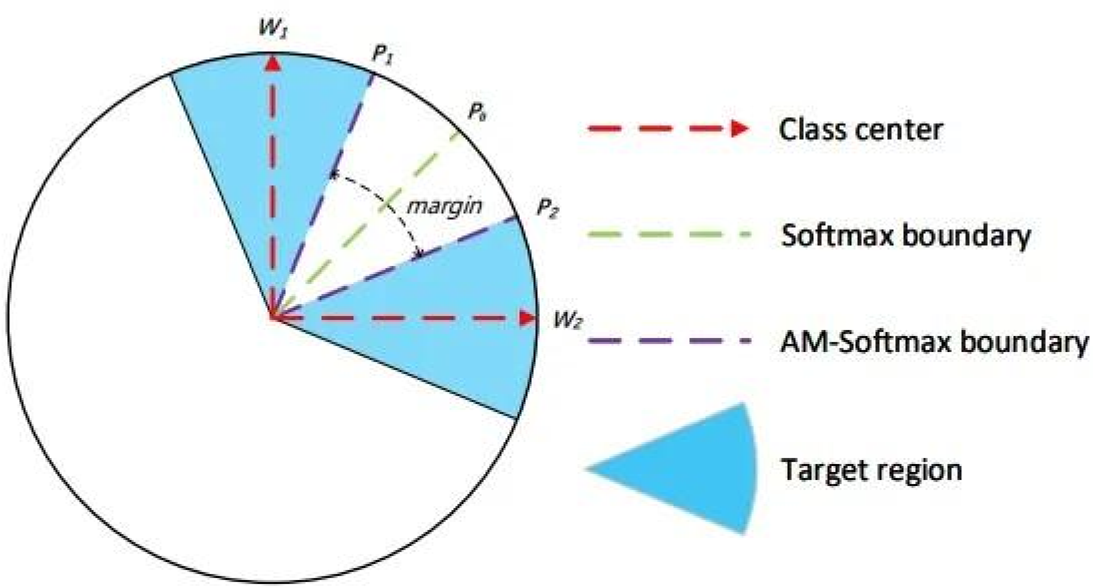
$$\psi(\theta) = cos\theta - m$$

$$cos\theta_{y_i} = \frac{W_{y_i}^T f_i}{\left\|W_{y_i}^T\right\| \left\|y_i\right\|} = W_{y_i}^T f_i$$

$$L_{AM-Softmax} = -\frac{1}{n}\sum_{i=1}^{n} log\frac{e^{s\cdot(cos\theta_{y_i}-m)}}{e^{s\cdot(cos\theta_{y_i}-m)}+\sum_{j=1,j\neq y_i}^{c} e^{s\cdot cos\theta_j}}$$
$$= -\frac{1}{n}\sum_{i=1}^{n} log\frac{e^{s\cdot(W_{y_i}^T f_i-m)}}{e^{s\cdot(W_{y_i}^T f_i-m)}+\sum_{j=1,j\neq y_i}^{c} e^{s\cdot W_j^T f_i}}$$

Where:

- n is the number of classes,
- s is a scaling hyper-parameter,
- m is the additive margin,
- theta_yi is the angle between the weight and feature of class (y_i),
- theta_j is the angle between the weight and feature of class (j).



APPLICATIONS:

In face recognition systems, the AM-Softmax loss function is designed to shrink within-class variation by putting emphasis on target logits, which in turn improves margin between target and non-target classes . This encourages the model to not only maximize the inter-class differences but also minimize the intra-class variations. This results in a model that can more effectively recognize different images of the same individual, leading to improved performance on face verification tasks .

In speaker verification, the AM-Softmax loss function has delivered remarkable performance by improving the discriminative power of deep neural networks . The AM-Softmax loss function introduces an additive margin into the softmax loss, which encourages the model to learn more discriminative speaker embeddings. This results in a model that can more effectively distinguish between different speakers, leading to improved performance on speaker verification tasks .

The AM-Softmax loss function has also been used in other domains such as expression recognition . In this domain, the AM-Softmax loss function has been shown to improve the accuracy of face expression recognition tasks based on convolutional neural networks .