

1. What parts does the author divide deep learning into? Are there any parts missing?
2. Why is the difference between sigmoid cross entropy and softmax cross entropy loss function?
3. What are some advantages of the absolute loss function over the squared loss function?

1. Object Detection and Face Recognition. The author does not mention Language models and other networks used in NLP
2. The sigmoid function is replaced by the softmax function otherwise.
3. Absolute loss does not increase rapidly with increase of error.