

1. What is the benefit of the loss function smoothing out when it approaches the global minimum (eg. the square loss, the Huber loss)
2. In what way the square loss function in regression is more sensitive to outliers compared to the absolute loss function
3. A candidate image for object detection tends to have more background regions than foreground regions (class imbalance). What can we add to the sigmoid cross entropy loss function to encourage the deep learning model to focus more on classifying the foreground regions, rather than the background?

1.

This makes sure the gradient update slowly converges to an optimum instead of making a big jump that misses the global minimum.

2.

The squared loss function squares the error term. When the error is larger, the loss increases faster, thus the squared loss function is more sensitive to outliers

3.

By introducing the focal loss weight to the cross-entropy function. The focal loss weight deemphasizes the loss of easy samples and emphasizes the loss of hard samples. The cross-entropy function with focal loss is written as

$$L = -\tilde{\tau}(1 - \tilde{p})^\gamma \log(\tilde{p})$$

When the input sample contains mostly background, it is easier for the model to classify, thus p will be large, and the focal loss weight $(1 - p)^\gamma$ thus will be small

When the input sample contains complicated foreground objects, it is harder for the model to classify, thus p will be small, the focal loss weight $(1 - p)^\gamma$ thus will be large.