

Quadratic

$$L_2 = \frac{1}{2} (y - f(x))^2$$

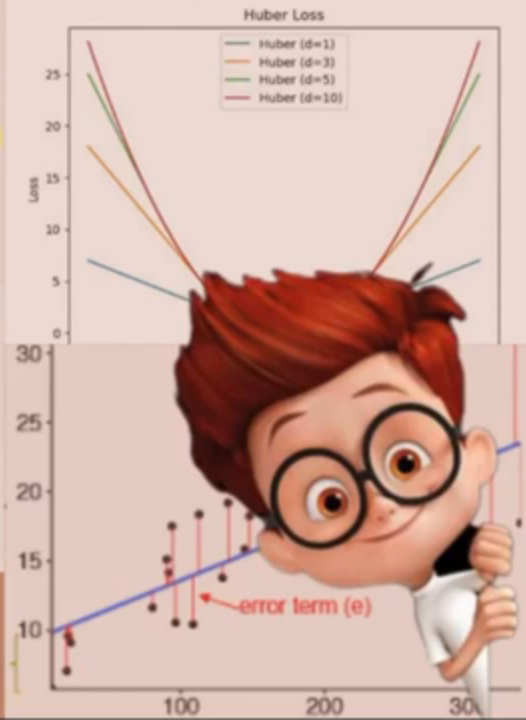
Smooth-L1 Huber

(4)

$$\frac{dL_1(x)}{dx} = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

(5)

Regression Losses



Quadratic

$$L_2 = \frac{1}{2} (y - f(x))^2$$

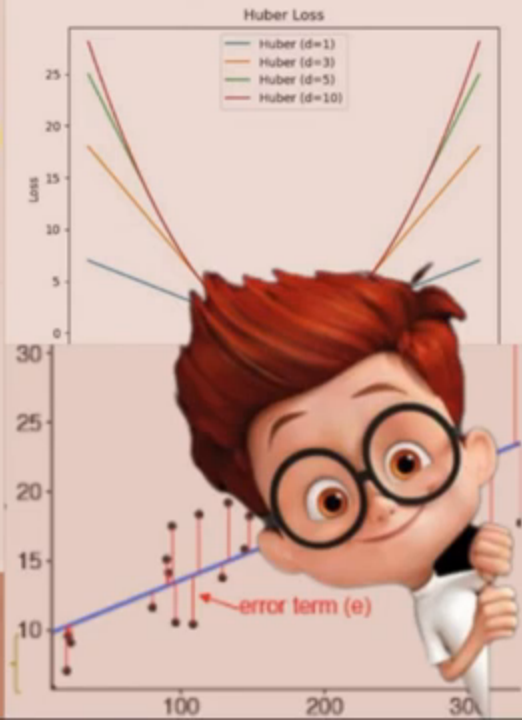
Smooth-L1 Huber

(4)

$$\frac{dL_1(x)}{dx} = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

(5)

Regression Losses



Quadratic

$$L_2 = \frac{1}{2} (y - f(x))^2$$

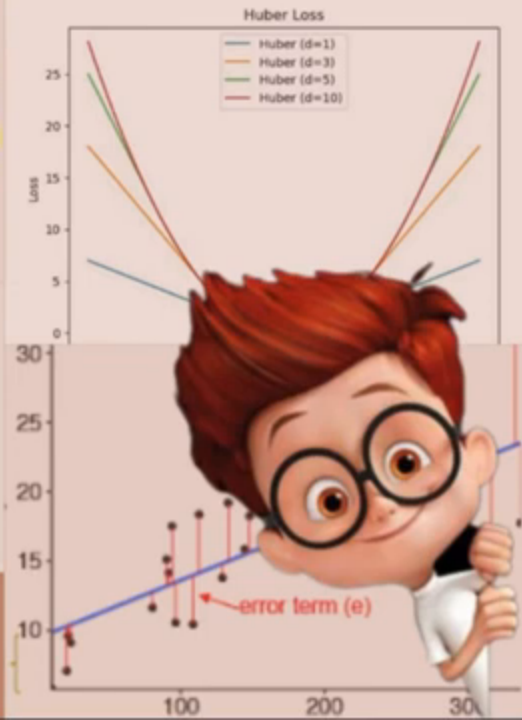
Smooth-L1 Huber

(4)

$$\frac{dL_1(x)}{dx} = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

(5)

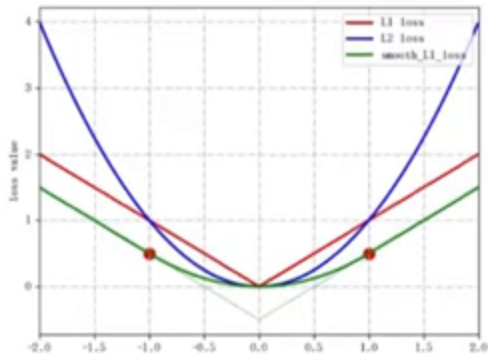
Regression Losses



Smooth L1 loss

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$

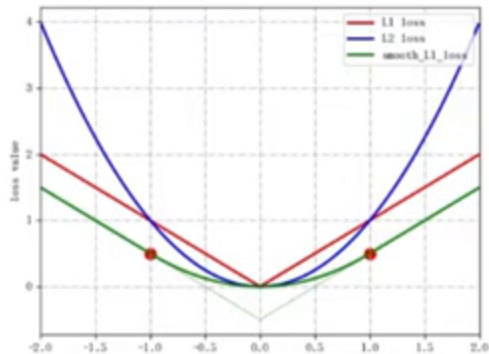
where $x = (Y^{\wedge} - Y)$
 Y^{\wedge} - Prediction
 Y - Target



Smooth L1 loss

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$

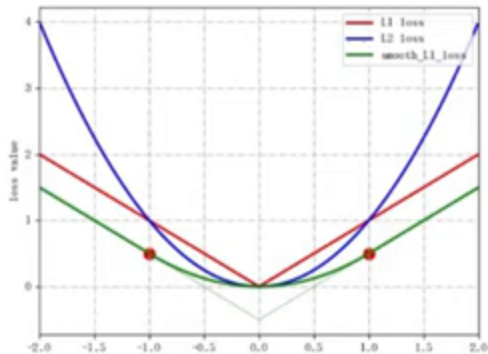
where $x = (Y^{\wedge} - Y)$
 Y^{\wedge} - Prediction
 Y - Target



Smooth L1 loss

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$

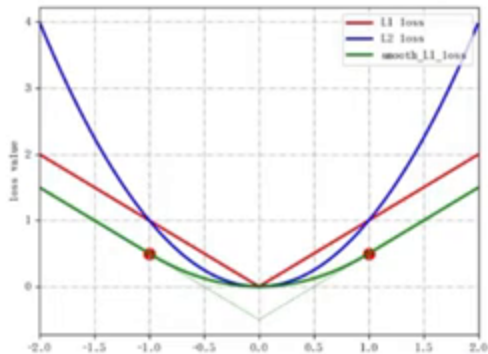
where $x = (Y^{\wedge} - Y)$
 Y^{\wedge} - Prediction
 Y - Target



Smooth L1 loss

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$

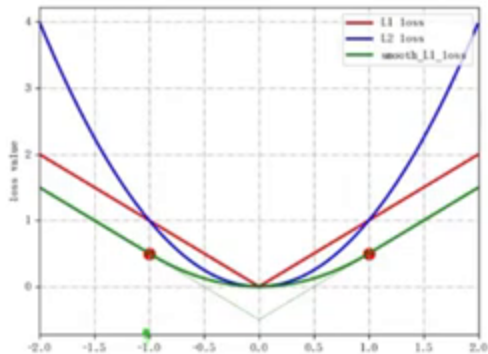
where $x = (\hat{Y} - Y)$
 \hat{Y} - Prediction
 Y - Target



Smooth L1 loss

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$

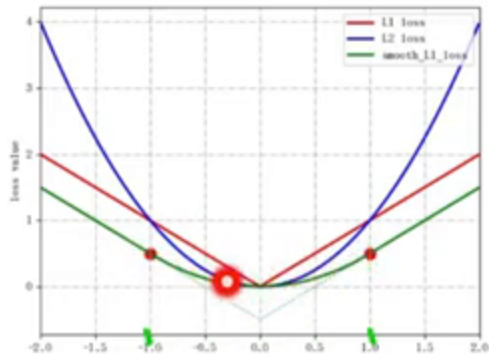
where $x = (\hat{Y} - Y)$
 \hat{Y} - Prediction
 Y - Target



Smooth L1 loss

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$

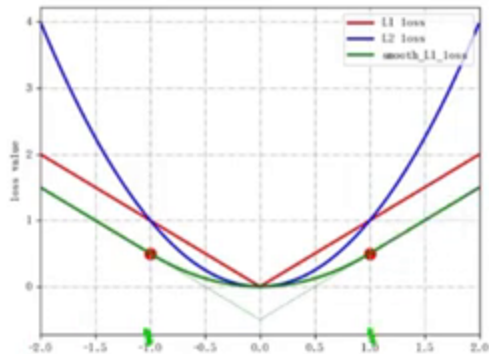
where $x = (\hat{Y} - Y)$
 \hat{Y} - Prediction
 Y - Target



Smooth L1 loss

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$

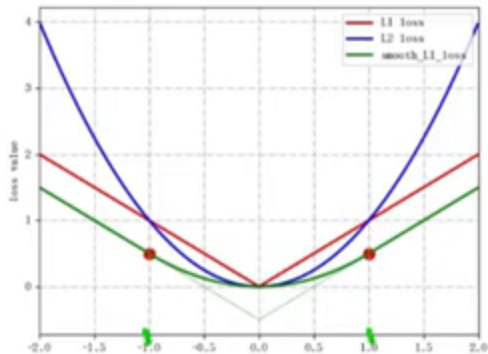
where $x = (\hat{Y} - Y)$
 \hat{Y} - Prediction
 Y - Target



Smooth L1 loss

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$

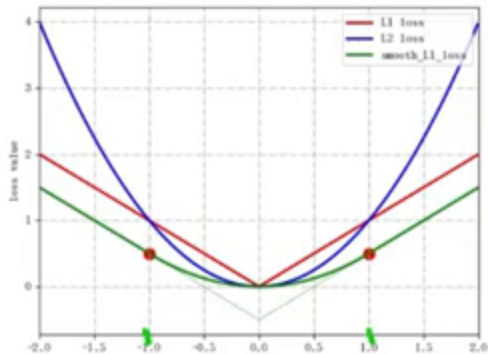
where $x = (\hat{Y} - Y)$
 \hat{Y} - Prediction
 Y - Target



Smooth L1 loss

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$

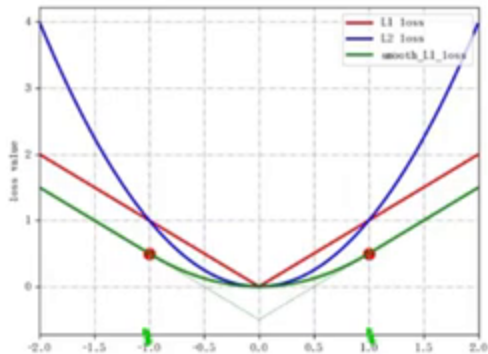
where $x = (\hat{Y} - Y)$
 \hat{Y} - Prediction
 Y - Target



Smooth L1 loss

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$

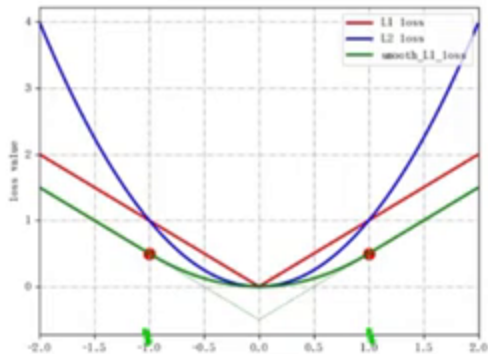
where $x = (\hat{Y} - Y)$
 \hat{Y} - Prediction
 Y - Target



Smooth L1 loss

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$

where $x = (\hat{Y} - Y)$
 \hat{Y} - Prediction
 Y - Target



Advantages

- Advantages of both L1 loss and L2 loss
- Less sensitive to outliers than L2 loss
- Prevents Exploding Gradients
- Well suited for most regression tasks

Advantages

- Advantages of both L1 loss and L2 loss
- Less sensitive to outliers than L2 loss
- Prevents Exploding Gradients
- Well suited for most regression tasks

Advantages

- Advantages of both L1 loss and L2 loss
- Less sensitive to outliers than L2 loss
- Prevents Exploding Gradients
- Well suited for most regression tasks

Advantages

- Advantages of both L1 loss and L2 loss
- Less sensitive to outliers than L2 loss
- Prevents Exploding Gradients
- Well suited for most regression tasks

Advantages

- Advantages of both L1 loss and L2 loss
- Less sensitive to outliers than L2 loss
- Prevents Exploding Gradients
- Well suited for most regression tasks

Advantages

- Advantages of both L1 loss and L2 loss
- Less sensitive to outliers than L2 loss
- Prevents Exploding Gradients
- Well suited for most regression tasks

Advantages

- Advantages of both L1 loss and L2 loss
- Less sensitive to outliers than L2 loss
- Prevents Exploding Gradients
- Well suited for most regression tasks

Derivative

$$L_2(x) = x^2$$

$$L_1(x) = |x|$$

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases}$$

Derivative

$$L_2(x) = x^2$$

$$L_1(x) = \underline{|x|}$$

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases}$$

$$\frac{dL_2(x)}{dx} = 2\underline{x}$$

$$\frac{dL_1(x)}{dx} = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

Derivative

$$L_2(x) = x^2$$

$$L_1(x) = \underline{|x|}$$

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases}$$

$$\frac{dL_2(x)}{dx} = 2\underline{x}$$

$$\frac{dL_1(x)}{dx} = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

$$\frac{d \text{smooth}_{L_1}}{dx} = \begin{cases} \underline{x} & \text{if } |x| < \underline{1} \\ \pm 1 & \text{otherwise} \end{cases}$$

Derivative

$$L_2(x) = x^2$$

$$L_1(x) = \underline{|x|}$$

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ \underline{|x|} - 0.5 & \text{otherwise} \end{cases}$$

$$\frac{dL_2(x)}{dx} = 2x$$

$$\frac{dL_1(x)}{dx} = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

$$\frac{d \text{smooth}_{L_1}}{dx} = \begin{cases} \pm x & \text{if } |x| < 1 \\ \pm 1 & \text{otherwise} \end{cases}$$

Derivative

$$L_2(x) = x^2$$

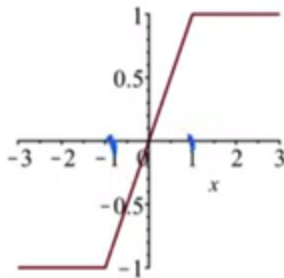
$$L_1(x) = \underline{|x|}$$

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ \underline{|x|} - 0.5 & \text{otherwise} \end{cases}$$

$$\frac{dL_2(x)}{dx} = 2x$$

$$\frac{dL_1(x)}{dx} = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

$$\frac{d \text{smooth}_{L_1}}{dx} = \begin{cases} \underline{x} & \text{if } |x| < 1 \\ \pm 1 & \text{otherwise} \end{cases}$$



Derivative

$$L_2(x) = x^2$$

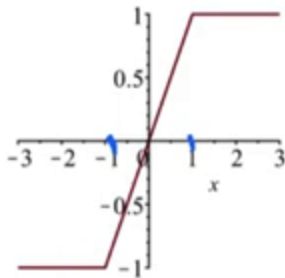
$$L_1(x) = \underline{|x|}$$

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ \underline{|x|} - 0.5 & \text{otherwise} \end{cases}$$

$$\frac{dL_2(x)}{dx} = 2x$$

$$\frac{dL_1(x)}{dx} = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

$$\frac{d \text{smooth}_{L_1}}{dx} = \begin{cases} \pm 1 & \text{if } |x| < 1 \\ \pm 1 & \text{otherwise} \end{cases}$$



Derivative

$$L_2(x) = x^2$$

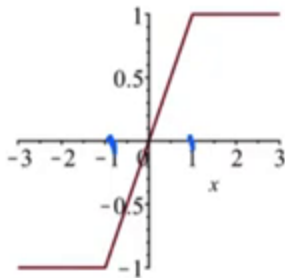
$$L_1(x) = \underline{|x|}$$

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ \underline{|x|} - 0.5 & \text{otherwise} \end{cases}$$

$$\frac{dL_2(x)}{dx} = 2x$$

$$\frac{dL_1(x)}{dx} = \begin{cases} 1 & \text{if } x \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

$$\frac{d \text{smooth}_{L_1}}{dx} = \begin{cases} \pm 1 & \text{if } |x| < 1 \\ \pm 1 & \text{otherwise} \end{cases}$$



Any Problem?

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$

Any Problem?

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$

Any Problem?

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$

Why 1?

Can I use 2 or 5?

Any Problem?

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$

Why 1?

Can I use 2 or 5?

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$



$$l_n = \begin{cases} 0.5(x_n - y_n)^2 / \text{beta}, & \text{if } |x_n - y_n| < \text{beta} \\ |x_n - y_n| - 0.5 * \text{beta}, & \text{otherwise} \end{cases}$$

Parameterized

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$



$$l_n = \begin{cases} 0.5(x_n - y_n)^2 / \text{beta}, & \text{if } |x_n - y_n| < \text{beta} \\ |x_n - y_n| - 0.5 * \text{beta}, & \text{otherwise} \end{cases}$$

Parameterized

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$



$$l_n = \begin{cases} 0.5(x_n - y_n)^2 / \text{beta}, & \text{if } |x_n - y_n| < \text{beta} \\ |x_n - y_n| - 0.5 * \text{beta}, & \text{otherwise} \end{cases}$$

Parameterized

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$



Any issues?

$$l_n = \begin{cases} 0.5(x_n - y_n)^2 / \text{beta}, & \text{if } |x_n - y_n| < \text{beta} \\ |x_n - y_n| - 0.5 * \text{beta}, & \text{otherwise} \end{cases}$$

Parameterized

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$



Any issues?

$$l_n = \begin{cases} 0.5(x_n - y_n)^2 / \text{beta}, & \text{if } |x_n - y_n| < \text{beta} \\ |x_n - y_n| - 0.5 * \text{beta}, & \text{otherwise} \end{cases}$$

Parameterized

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$



Any issues?

$$l_n = \begin{cases} 0.5(x_n - y_n)^2 / \text{beta}, & \text{if } |x_n - y_n| < \text{beta} \\ |x_n - y_n| - 0.5 * \text{beta}, & \text{otherwise} \end{cases}$$

Parameterized

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$



Any issues?

$$l_n = \begin{cases} 0.5(x_n - y_n)^2 / \text{beta}, & \text{if } |x_n - y_n| < \text{beta} \\ |x_n - y_n| - 0.5 * \text{beta}, & \text{otherwise} \end{cases}$$

Parameterized

$$\text{smooth}_{L_1}(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise,} \end{cases}$$



Any issues?

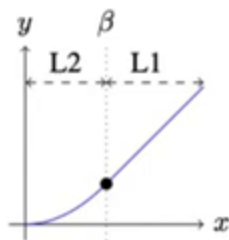
$$l_n = \begin{cases} 0.5(x_n - y_n)^2 / \text{beta}, & \text{if } |x_n - y_n| < \text{beta} \\ |x_n - y_n| - 0.5 * \text{beta}, & \text{otherwise} \end{cases}$$

Parameterized

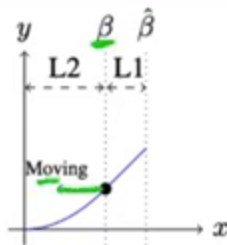
Dynamic SmoothL1 Loss

Dynamic SmoothL1 Loss

Self-Adjusting Smooth L1 Loss



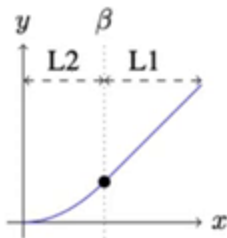
(a) Smooth L1



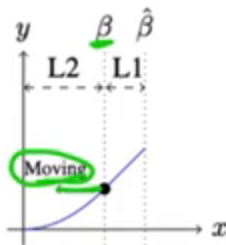
(b) Self-Adjusting Smooth L1

Dynamic SmoothL1 Loss

Self-Adjusting Smooth L1 Loss



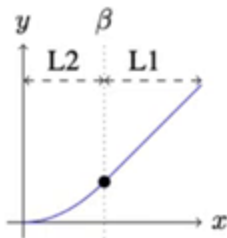
(a) Smooth L1



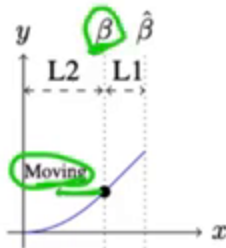
(b) Self-Adjusting Smooth L1

Dynamic SmoothL1 Loss

Self-Adjusting Smooth L1 Loss



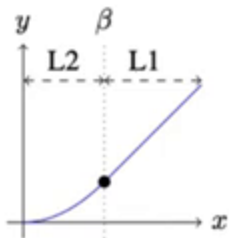
(a) Smooth L1



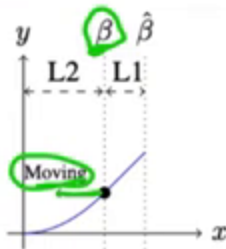
(b) Self-Adjusting Smooth L1

Dynamic SmoothL1 Loss

Self-Adjusting Smooth L1 Loss



(a) Smooth L1



(b) Self-Adjusting Smooth L1

Huber Loss

<https://github.com> > torch > issues > 1

Smooth L1 loss vs Huber loss · Issue #579 · torch/nn - GitHub

22-Jan-2016 — I would say that the Huber loss really is parameterised by delta, as it defines the boundary between the squared and absolute costs. I don't ...

Huber Loss

<https://github.com/torch/torch/issues>

Smooth L1 loss vs Huber loss · Issue #579 · torch/nn - GitHub

22-Jan-2016 — I would say that the Huber loss really is parameterised by delta, as it defines the boundary between the squared and absolute costs. I don't ...



TensorFlow

Install

TensorFlow v2.10.0

Overview

Python

C++

J

▼ Filter

KLDivergence

LogCosh

Loss

MeanAbsoluteError

MeanAbsolutePercentageError

MeanSquaredError

MeanSquaredLogarithmicError

Poisson

Reduction

SparseCategoricalCrossentropy

SquaredHinge

categorical_hinge

cosine_similarity

deserialize

get

huber

log_cosh

serialize

► metrics

Huber Loss

$$l_n = \begin{cases} 0.5(x_n - y_n)^2 / \textit{beta}, & \text{if } |x_n - y_n| < \textit{beta} \\ |x_n - y_n| - 0.5 * \textit{beta}, & \text{otherwise} \end{cases}$$

Smooth-L1

$$l_n = \begin{cases} 0.5(x_n - y_n)^2, & \text{if } |x_n - y_n| < \textit{delta} \\ \textit{delta} * (|x_n - y_n| - 0.5 * \textit{delta}), & \text{otherwise} \end{cases}$$

Huber

Huber Loss

$$l_n = \begin{cases} 0.5(x_n - y_n)^2 / \textit{beta}, & \text{if } |x_n - y_n| < \textit{beta} \\ |x_n - y_n| - 0.5 * \textit{beta}, & \text{otherwise} \end{cases}$$

Smooth-L1

$$l_n = \begin{cases} 0.5(x_n - y_n)^2, & \text{if } |x_n - y_n| < \textit{delta} \\ \textit{delta} * (|x_n - y_n| - 0.5 * \textit{delta}), & \text{otherwise} \end{cases}$$

Huber

Huber Loss


$$l_n = \begin{cases} 0.5(x_n - y_n)^2 / \textit{beta}, & \text{if } |x_n - y_n| < \textit{beta} \\ |x_n - y_n| - 0.5 * \textit{beta}, & \text{otherwise} \end{cases}$$

Smooth-L1

$$l_n = \begin{cases} 0.5(x_n - y_n)^2, & \text{if } |x_n - y_n| < \textit{delta} \\ \textit{delta} * (|x_n - y_n| - 0.5 * \textit{delta}), & \text{otherwise} \end{cases}$$

Huber

Huber Loss

$$l_n = \begin{cases} 0.5(x_n - y_n)^2 / \textit{beta}, & \text{if } |x_n - y_n| < \textit{beta} \\ |x_n - y_n| - 0.5 * \textit{beta}, & \text{otherwise} \end{cases}$$


Smooth-L1

$$l_n = \begin{cases} 0.5(x_n - y_n)^2, & \text{if } |x_n - y_n| < \textit{delta} \\ \textit{delta} * (|x_n - y_n| - 0.5 * \textit{delta}), & \text{otherwise} \end{cases}$$

Huber

Huber Loss

$$l_n = \begin{cases} 0.5(x_n - y_n)^2 / \text{beta}, & \text{if } |x_n - y_n| < \text{beta} \\ |x_n - y_n| - 0.5 * \text{beta}, & \text{otherwise} \end{cases}$$

Smooth-L1

beta : delta

$$l_n = \begin{cases} 0.5(x_n - y_n)^2, & \text{if } |x_n - y_n| < \text{delta} \\ \text{delta} * (|x_n - y_n| - 0.5 * \text{delta}), & \text{otherwise} \end{cases}$$

Huber

Huber Loss

$$l_n = \begin{cases} 0.5(x_n - y_n)^2 / \textit{beta}, & \text{if } |x_n - y_n| < \textit{beta} \\ |x_n - y_n| - 0.5 * \textit{beta}, & \text{otherwise} \end{cases}$$

Smooth-L1

$$l_n = \begin{cases} 0.5(x_n - y_n)^2, & \text{if } |x_n - y_n| < \textit{delta} \\ \textit{delta} * (|x_n - y_n| - 0.5 * \textit{delta}), & \text{otherwise} \end{cases}$$

Huber

Huber Loss

$$l_n = \begin{cases} 0.5(x_n - y_n)^2 / \text{beta}, & \text{if } |x_n - y_n| < \text{beta} \\ |x_n - y_n| - 0.5 * \text{beta}, & \text{otherwise} \end{cases}$$

Smooth-L1

$$l_n = \begin{cases} 0.5(x_n - y_n)^2, & \text{if } |x_n - y_n| < \text{delta} \\ \text{delta} * (|x_n - y_n| - 0.5 * \text{delta}), & \text{otherwise} \end{cases}$$

Huber = beta (S-L1)

Huber Loss

$$l_n = \begin{cases} 0.5(x_n - y_n)^2 / \text{beta}, & \text{if } |x_n - y_n| < \text{beta} \\ |x_n - y_n| - 0.5 * \text{beta}, & \text{otherwise} \end{cases}$$

Smooth-L1

$$l_n = \begin{cases} 0.5(x_n - y_n)^2, & \text{if } |x_n - y_n| < \text{delta} \\ \text{delta} * (|x_n - y_n| - 0.5 * \text{delta}), & \text{otherwise} \end{cases}$$

Huber = poly (S-L1)

Huber Loss

$$l_n = \begin{cases} 0.5(x_n - y_n)^2 / \text{beta}, & \text{if } |x_n - y_n| < \text{beta} \\ |x_n - y_n| - 0.5 * \text{beta}, & \text{otherwise} \end{cases}$$

Smooth-L1

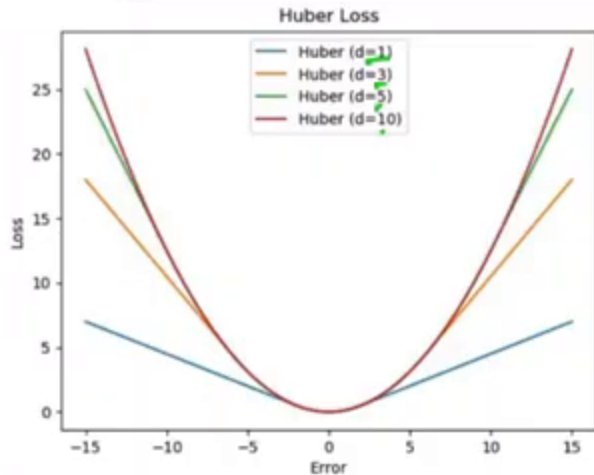
$$l_n = \begin{cases} 0.5(x_n - y_n)^2, & \text{if } |x_n - y_n| < \text{delta} \\ \text{delta} * (|x_n - y_n| - 0.5 * \text{delta}), & \text{otherwise} \end{cases}$$

Huber = beta (5-17)

3

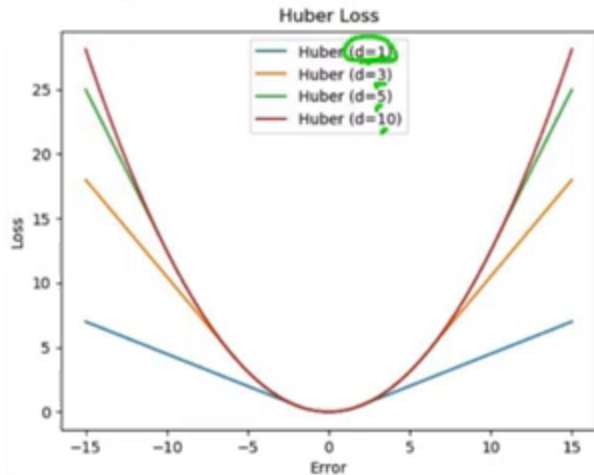
Huber Loss

$$l_n = \begin{cases} 0.5(x_n - y_n)^2, & \text{if } |x_n - y_n| < \text{delta} \\ \text{delta} * (|x_n - y_n| - 0.5 * \text{delta}), & \text{otherwise} \end{cases}$$



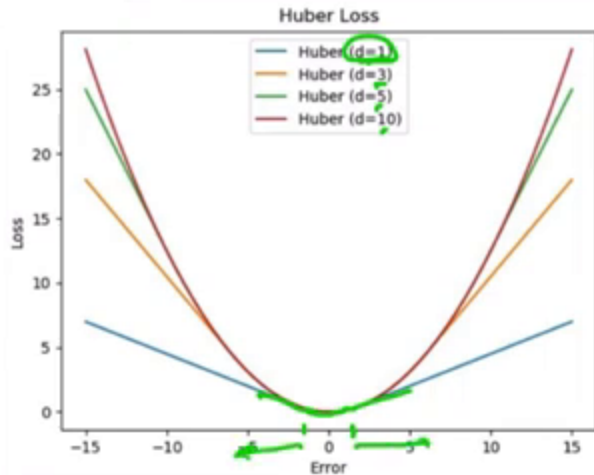
Huber Loss

$$l_n = \begin{cases} 0.5(x_n - y_n)^2, & \text{if } |x_n - y_n| < \text{delta} \\ \text{delta} * (|x_n - y_n| - 0.5 * \text{delta}), & \text{otherwise} \end{cases}$$



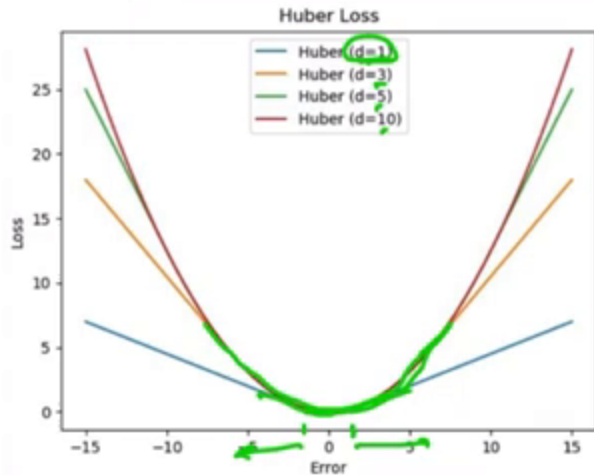
Huber Loss

$$l_n = \begin{cases} 0.5(x_n - y_n)^2, & \text{if } |x_n - y_n| < \text{delta} \\ \text{delta} * (|x_n - y_n| - 0.5 * \text{delta}), & \text{otherwise} \end{cases}$$



Huber Loss

$$l_n = \begin{cases} 0.5(x_n - y_n)^2, & \text{if } |x_n - y_n| < \text{delta} \\ \text{delta} * (|x_n - y_n| - 0.5 * \text{delta}), & \text{otherwise} \end{cases}$$



Implementation

```
def smoothl1loss(x, y):  
    if abs(x-y) < 1: return 1/2*(x-y)**2  
    else: return abs(x-y)-1/2
```