Круглов А.И. 605-202

$\boxed{4}$ Запишем Huber Loss в немного более удобном виде:

$$L(\theta) = \begin{cases} \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{2}(y_i - x_i^T\theta)^2\right), & y_i - x_i^T\theta < c \\ \frac{1}{n}\sum_{i=1}^{n}\left(c\,|y_i - x_i^T\theta| - \frac{c^2}{2}\right), & y_i - x_i^T\theta \geqslant c \end{cases}$$

* перед $\Sigma$ должен быть множитель $\frac{1}{n}$, видимо, в условии опечатка

$$\nabla L = \begin{cases} \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i^T\theta)(-x_i), & y_i - x_i^T\theta < c \\ \frac{1}{n}\sum_{i=1}^{n} c\cdot \operatorname{sign}(y_i - x_i^T\theta)(-x_i), & y_i - x_i^T\theta \geqslant c \end{cases}$$

объединим всё в 1 выражение с помощью индикаторов

$$\nabla L = \frac{1}{n}\sum_{i=1}^{n}\left((y_i - x_i^T\theta)\cdot \mathbb{1}\{y_i - x_i^T\theta < c\} + c\cdot\operatorname{sign}(y_i - x_i^T\theta)\cdot\mathbb{1}\{y_i - x_i^T\theta \geqslant c\}\right)\cdot(-x_i)$$

GD: $\theta_{k+1} = \theta_k - \eta\cdot\nabla L(\theta_k) = \theta_k - \frac{\eta}{n}\sum_{i=1}^{n}\left((y_i - x_i^T\theta)\cdot\mathbb{1}_< + c\cdot\operatorname{sign}(y_i - x_i^T\theta)\cdot\mathbb{1}_\geqslant\right)\cdot(-x_i)$

SGD: $\theta_{k+1} = \theta_k - \frac{\eta}{p}\sum\left((y_i - x_i^T\theta)\cdot\mathbb{1}_< + c\cdot\operatorname{sign}(y_i - x_i^T\theta)\cdot\mathbb{1}_\geqslant\right)\cdot(-x_i)$, где сумма $\Sigma$ идет по батчу размера $p$

Польза Huber Loss в том, что она совмещает плюсы L1 и L2 лоссов. Вблизи нуля она ведет себя как MSE (дифференцируемая, гладкая). На бесконечности она ведет себя как MAE (растет линейно, т.е. медленнее MSE, а значит, менее чувствительна к выбросам)