COMP2261 ARTIFICIAL INTELLIGENCE / MACHINE LEARNING

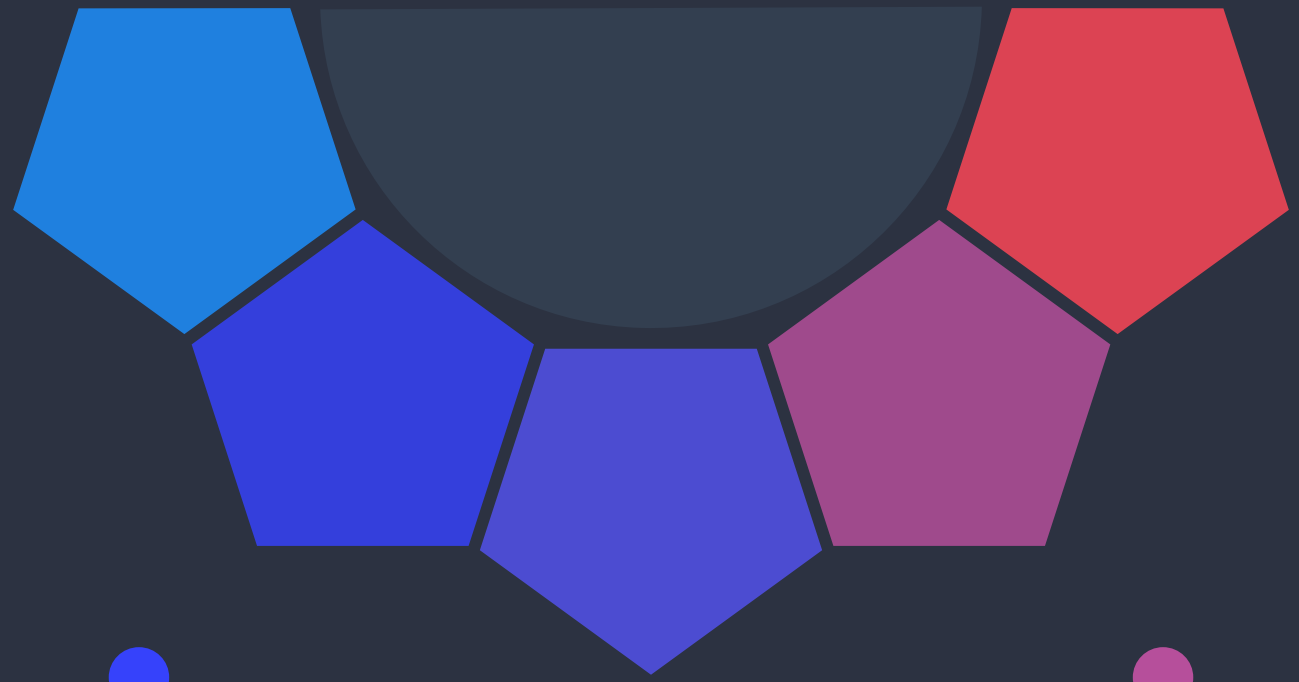
Cost Functions
For Regression Models
-- Huber-M

Dr SHI Lei

Durham University
learning lab

Learning Objectives

- Understand what is Huber-M
- Understand how Huber-M cost function works
- Understand the hyperparameter δ



MSE

Quantile

MAE

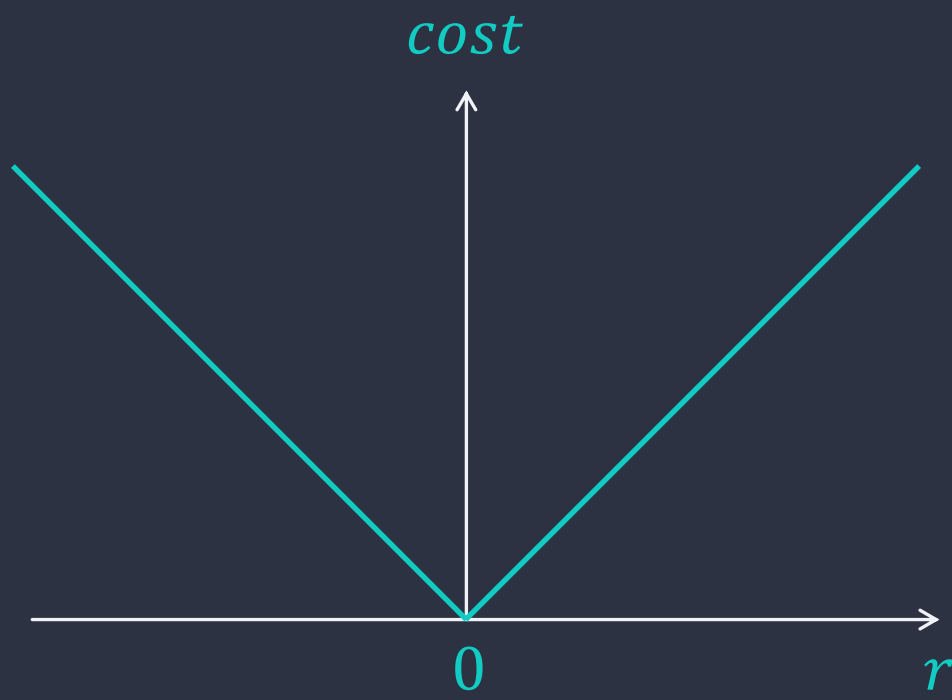
Log-Cosh

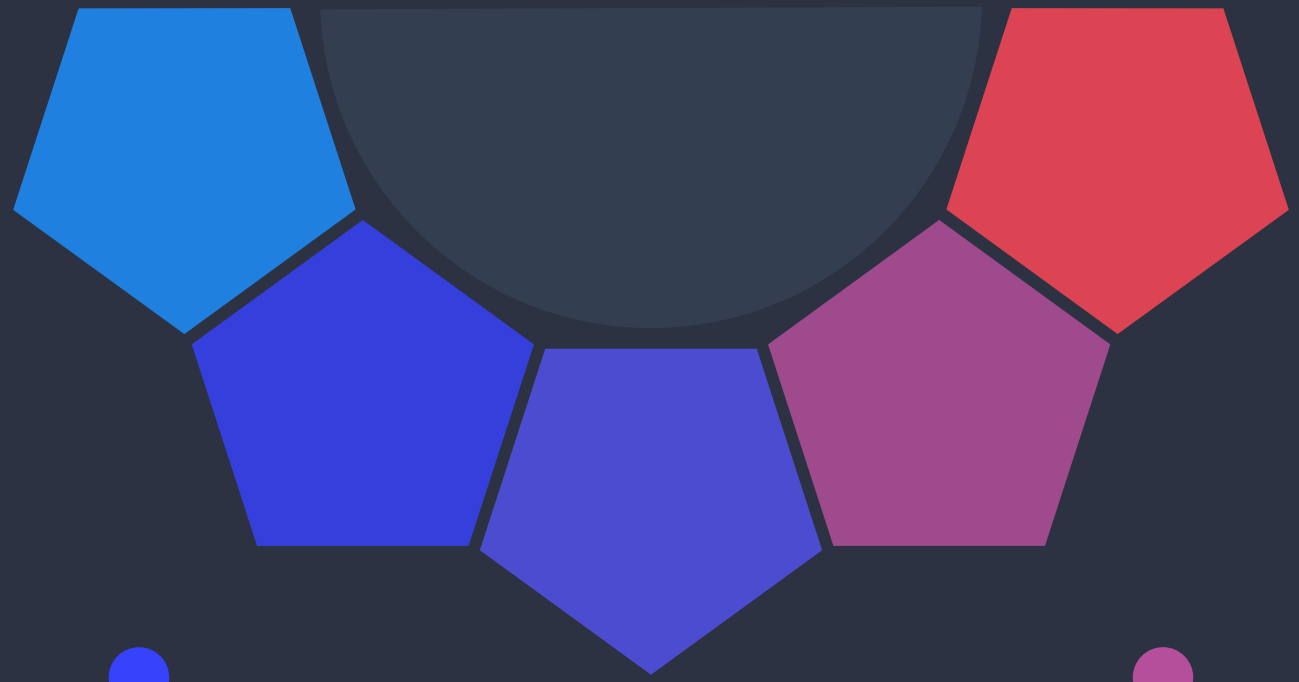
Huber-M

Mean Squared Error (MSE) vs Mean Absolute Error (MAE)

$$J = \frac{1}{m}\sum_{i=1}^{m}(y^{(i)} - \hat{y}^{(i)})^2$$

$$J = \frac{1}{m}\sum_{i=1}^{m}|y^{(i)} - \hat{y}^{(i)}|$$





MSE

Quantile

MAE

Log-Cosh

Huber-M

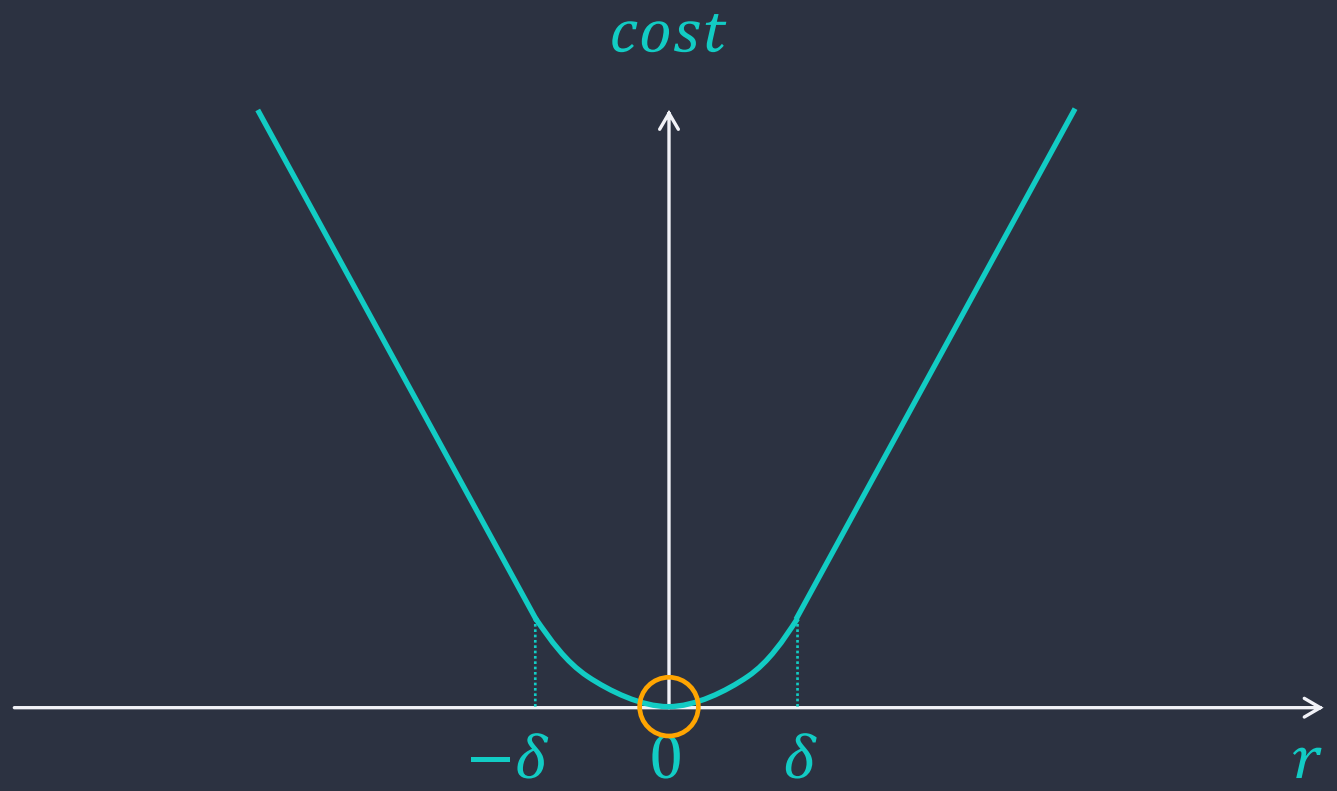
Huber-M

Huber-M

$$J = \frac{1}{m}\sum_{i=1}^{m}\begin{cases}\frac{1}{2}(y^{(i)} - \hat{y}^{(i)})^2, & \left|y^{(i)} - \hat{y}^{(i)}\right| \le \delta \\ \delta\left(\left|y^{(i)} - \hat{y}^{(i)}\right| - \frac{1}{2}\delta\right), & otherwise\end{cases}$$

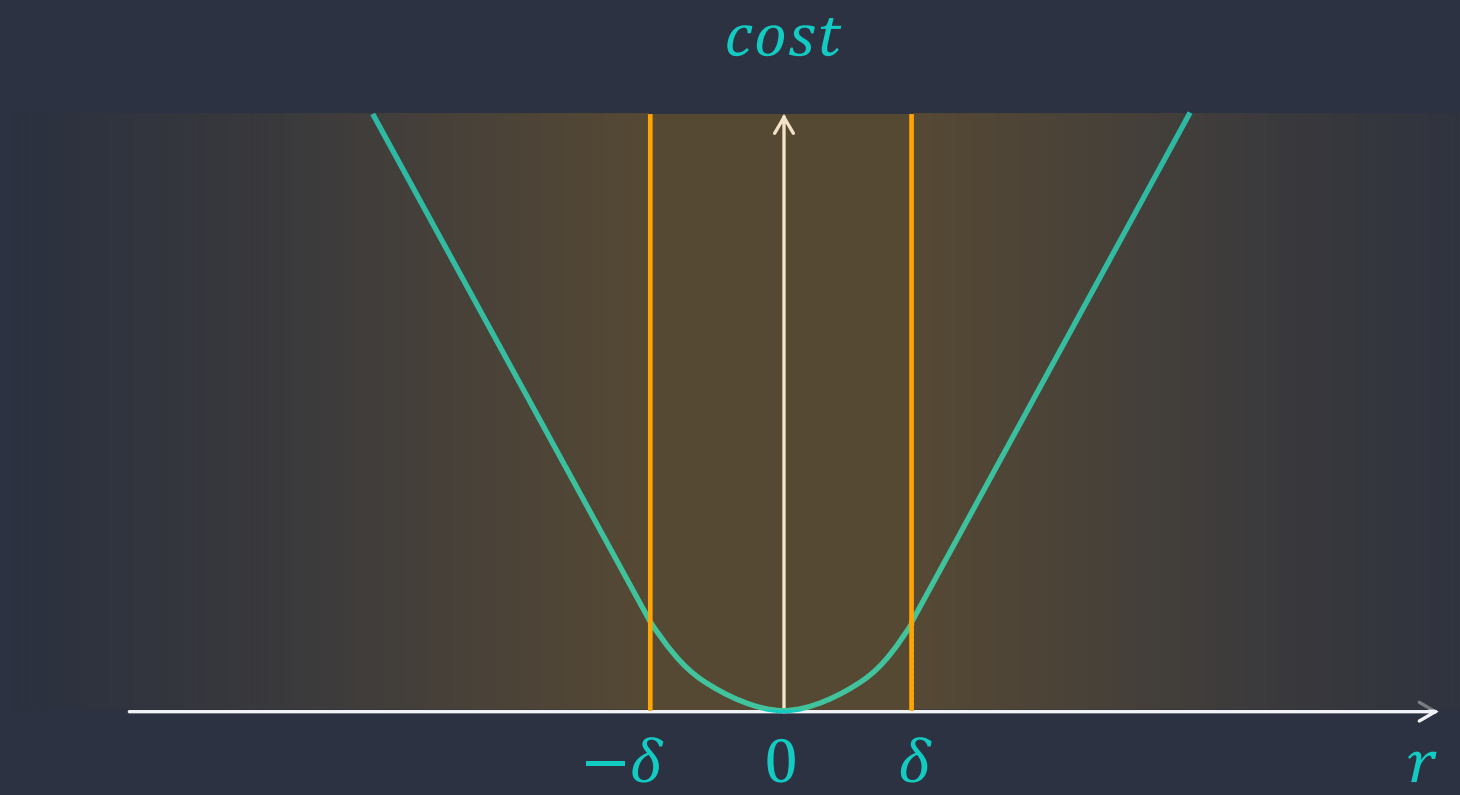
Huber-M

$$J = \frac{1}{m}\sum_{i=1}^{m}\begin{cases}\frac{1}{2}(y^{(i)} - \hat{y}^{(i)})^2, & \left|y^{(i)} - \hat{y}^{(i)}\right| \le \delta \\ \delta\left(\left|y^{(i)} - \hat{y}^{(i)}\right| - \frac{1}{2}\delta\right), & otherwise\end{cases}$$



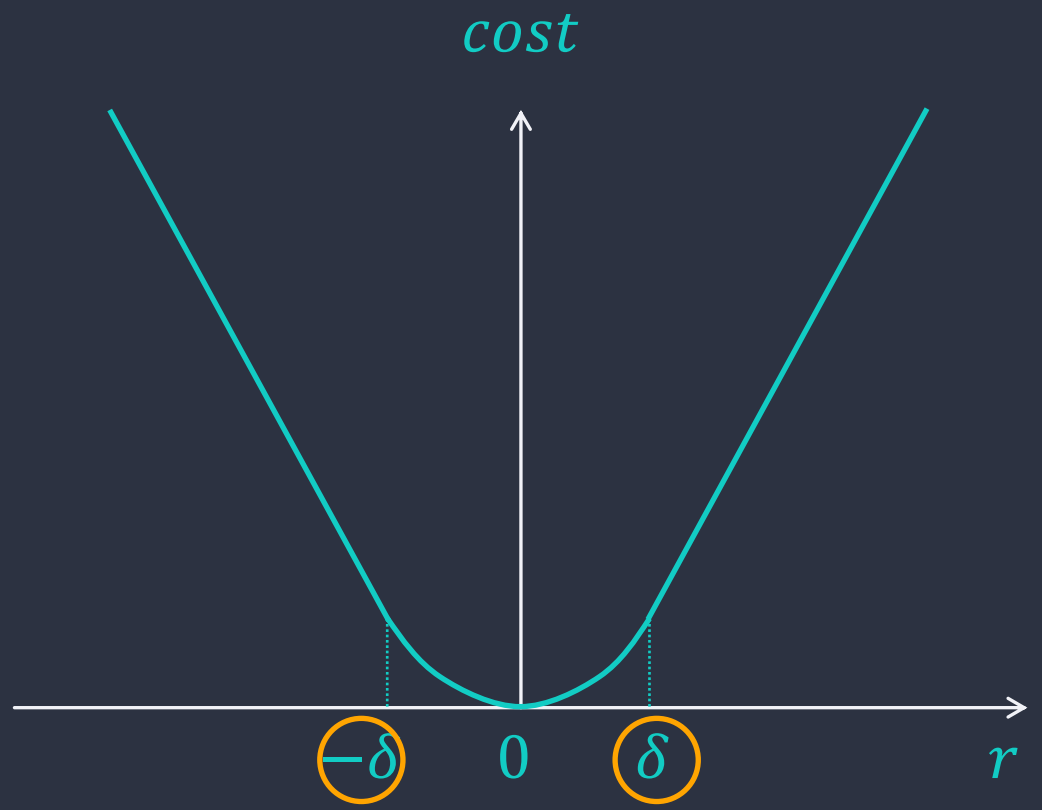
Huber-M

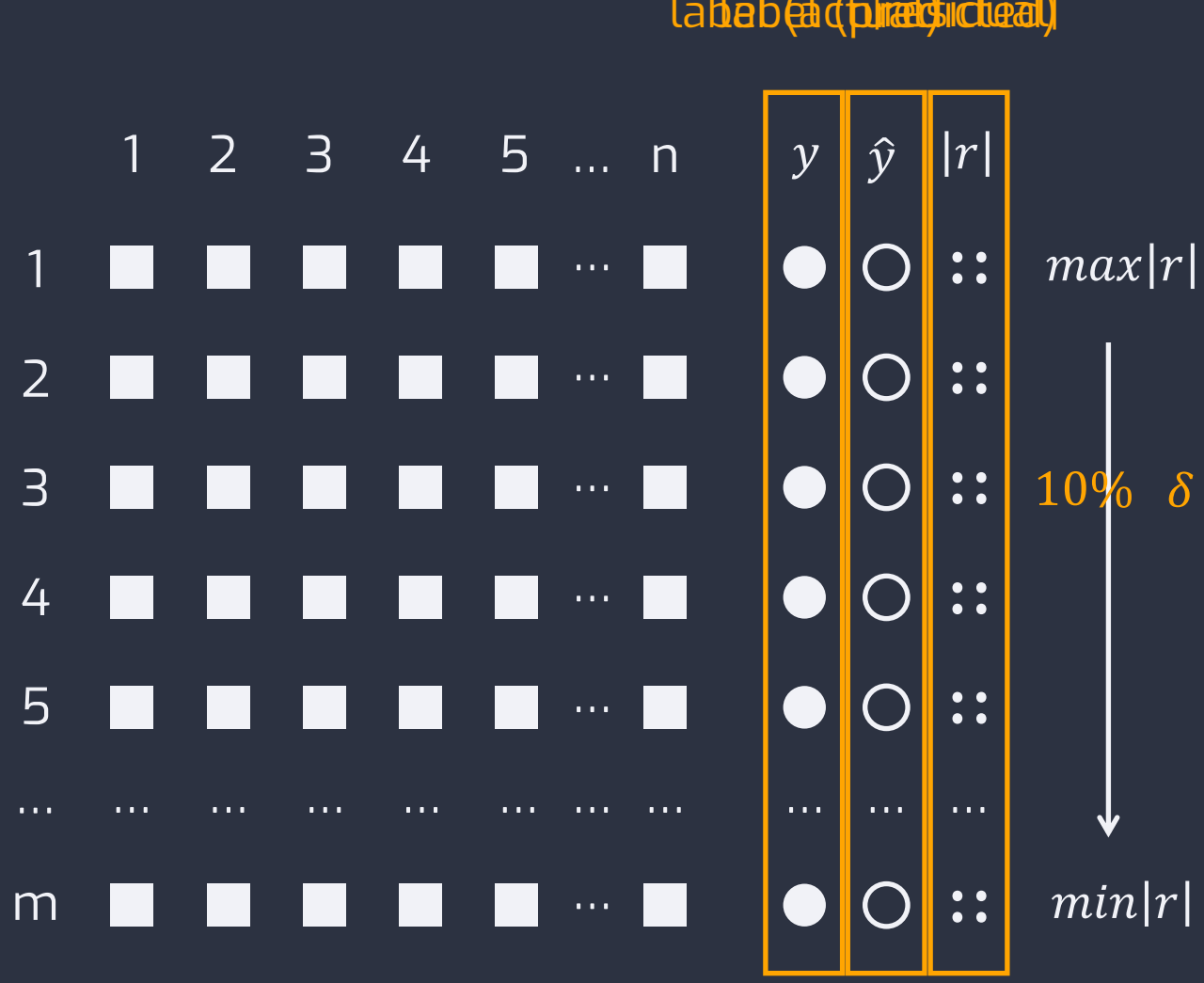
$$J = \frac{1}{m}\sum_{i=1}^{m}\begin{cases}\frac{1}{2}(y^{(i)} - \hat{y}^{(i)})^2, & \left|y^{(i)} - \hat{y}^{(i)}\right| \le \delta \\ \delta\left(\left|y^{(i)} - \hat{y}^{(i)}\right| - \frac{1}{2}\delta\right), & otherwise\end{cases}$$



Huber-M

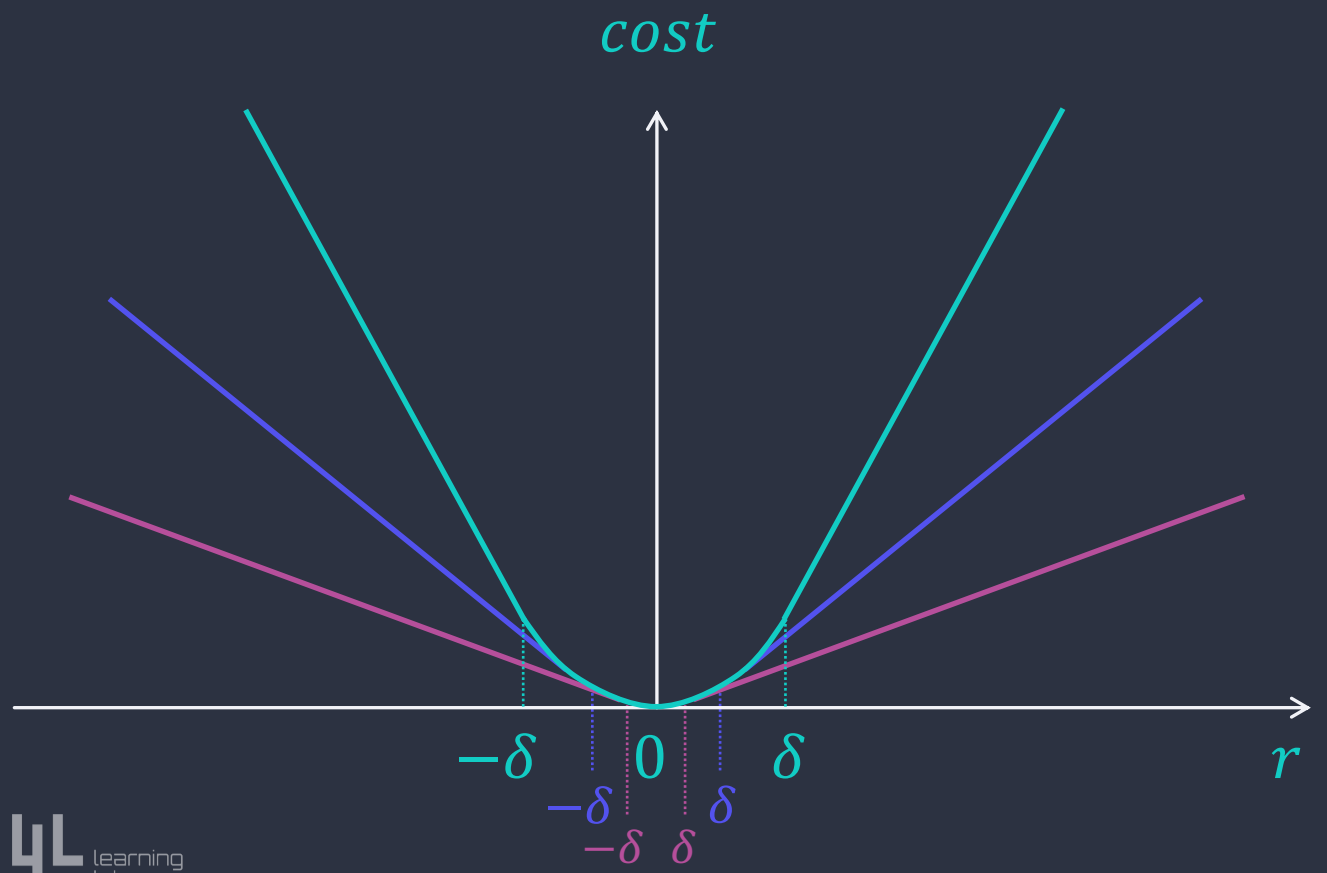
$$J = \frac{1}{m}\sum_{i=1}^{m}\begin{cases}\frac{1}{2}(y^{(i)} - \hat{y}^{(i)})^2, & |y^{(i)} - \hat{y}^{(i)}| \le \delta \\ \delta(|y^{(i)} - \hat{y}^{(i)}| - \frac{1}{2}\delta), & otherwise\end{cases}$$





Huber-M

$$J = \frac{1}{m}\sum_{i=1}^{m}\begin{cases}\frac{1}{2}(y^{(i)} - \hat{y}^{(i)})^2, & \left|y^{(i)} - \hat{y}^{(i)}\right| \le \delta \\ \delta\left(\left|y^{(i)} - \hat{y}^{(i)}\right| - \frac{1}{2}\delta\right), & otherwise\end{cases}$$



Choice of δ is very important, as it determines what are considered as outliers and how they are delt with.

Takeaway Points

- Huber-M combines good properties from MSE and MAE.
- Curving around minima – reducing gradient during training.
- More robust with outliers.
- Important to choose hyperparameter δ as it determines what are considered as outliers and how to deal with them.