COMP2261 ARTIFICIAL INTELLIGENCE / MACHINE LEARNING

Cost Functions

For Regression Models

-- MSE & MAE

Dr SHI Lei

Durham University

learning lab

1. Problem Framing
2. Data Preparation
3. Model Selection
4. Model Training
5. Model Testing
6. Hyperparameter Tuning
7. Inference / Prediction

Cost Functions

Learning Objectives

- Understand pitfall of the Mean Squared Error cost function
- Understand the alternatives to the MSE cost function
- Understand the differences between MSE and MAE

Supervised Learning

- To build a model represented as a hypothesis function $h(x)$.

income (input x, dependent variable)

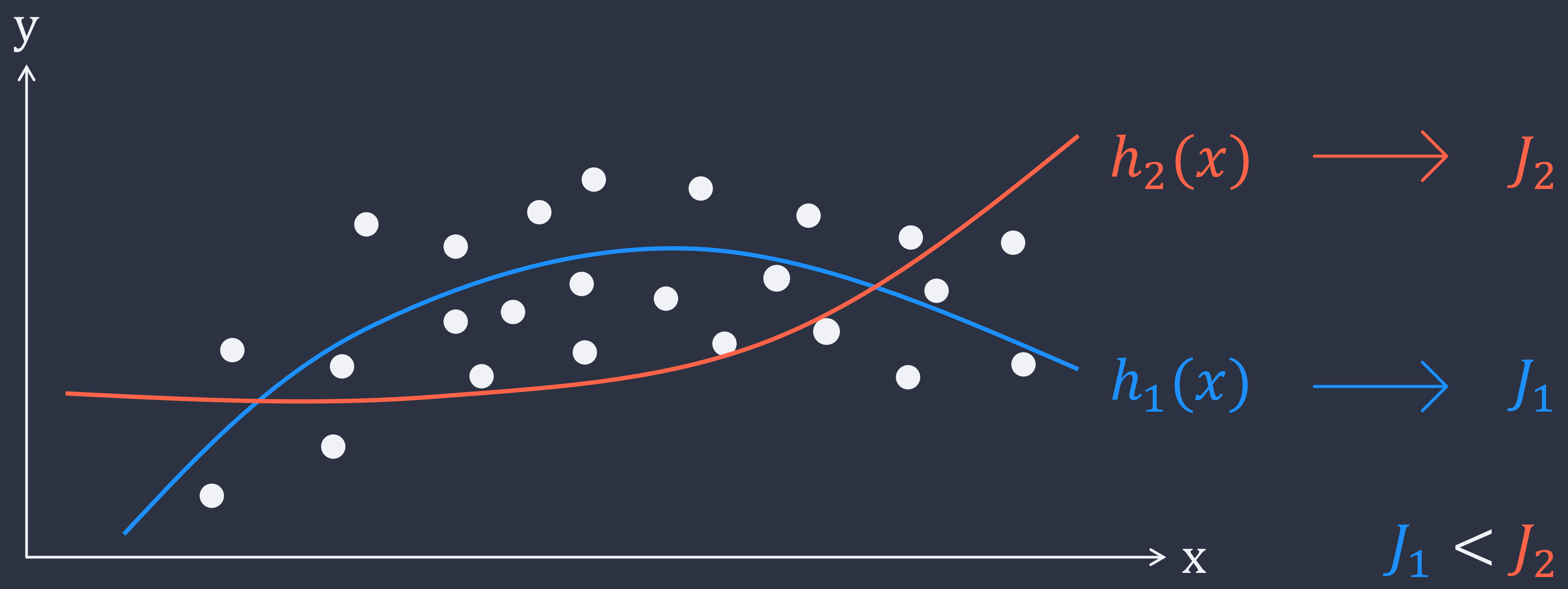
↓

model (hypothesis function, mapping $x \rightarrow y$) ⇦ cost function (loss function)

↓

happiness (output y, independent variable)

Cost Function for Regression



- The smaller values of the cost function, the better the model fits the dataset.
- Cost function to compare predicted values and actual values, using specific measure of "goodness of it".

MSE

Quantile

MAE

Log-Cosh

Huber-M

Mean Squared Error (MSE)

Mean Squared Error (MSE)

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

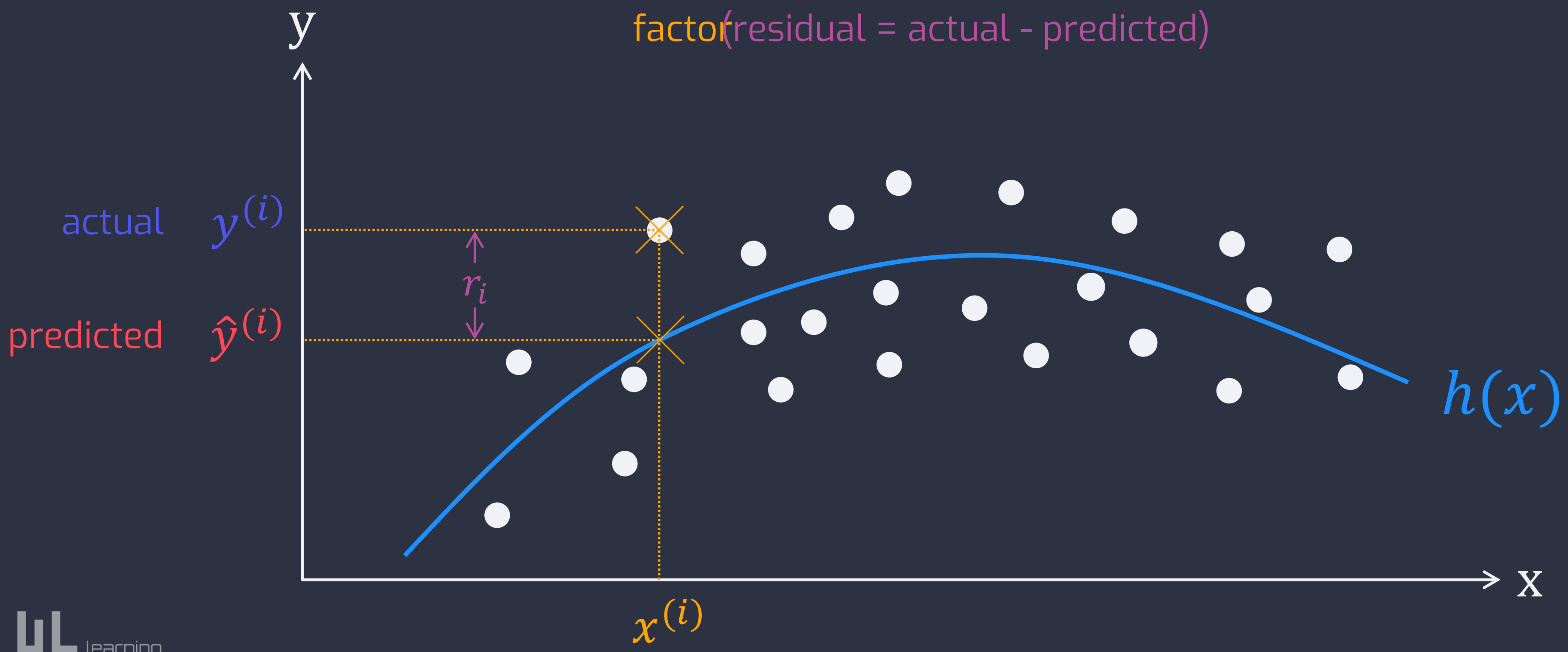
actual predicted

Mean Squared Error (MSE)

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

Normalising factor

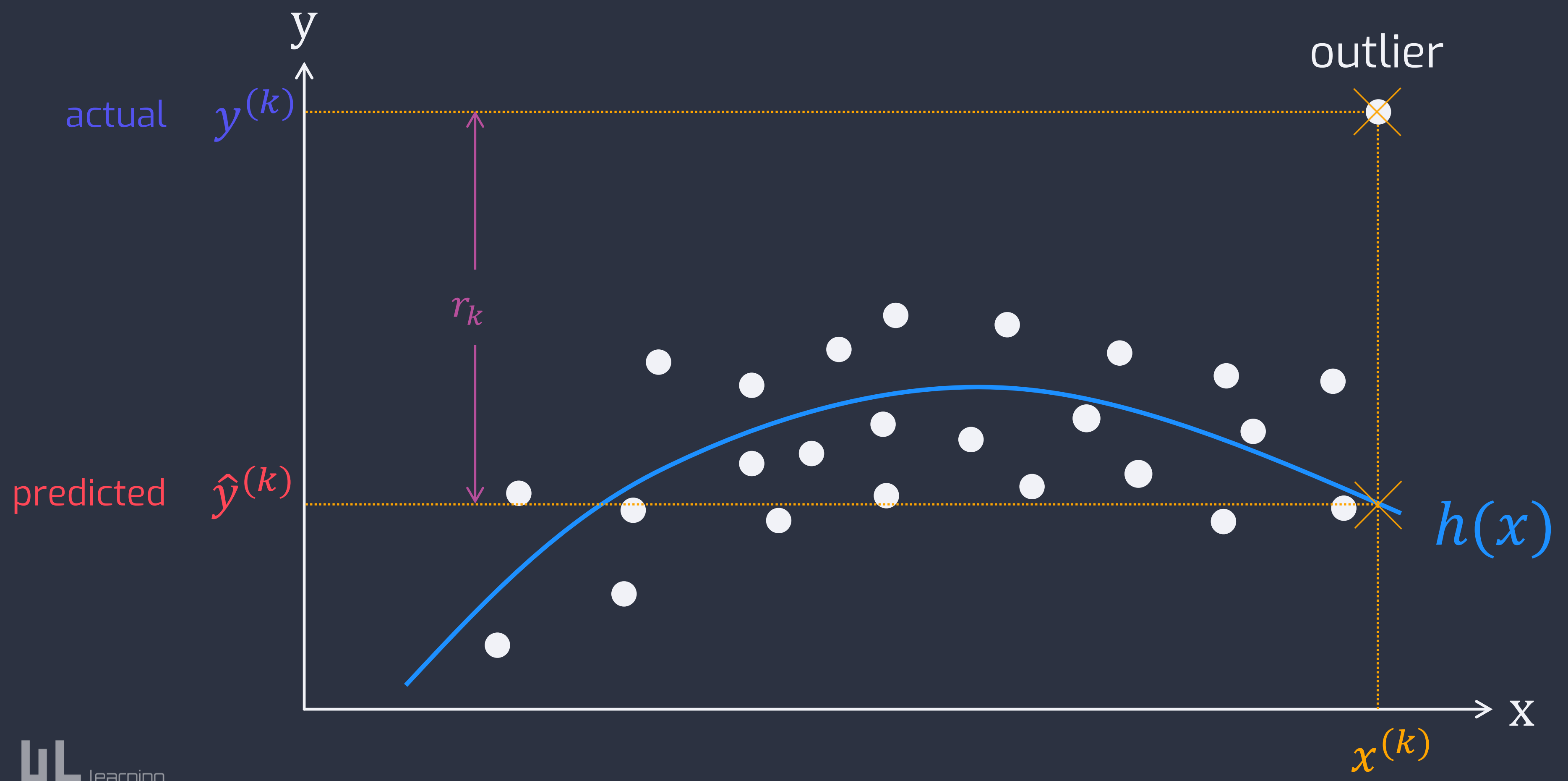
r_i (residual = actual - predicted)



Mean Squared Error (MSE) Cost Function

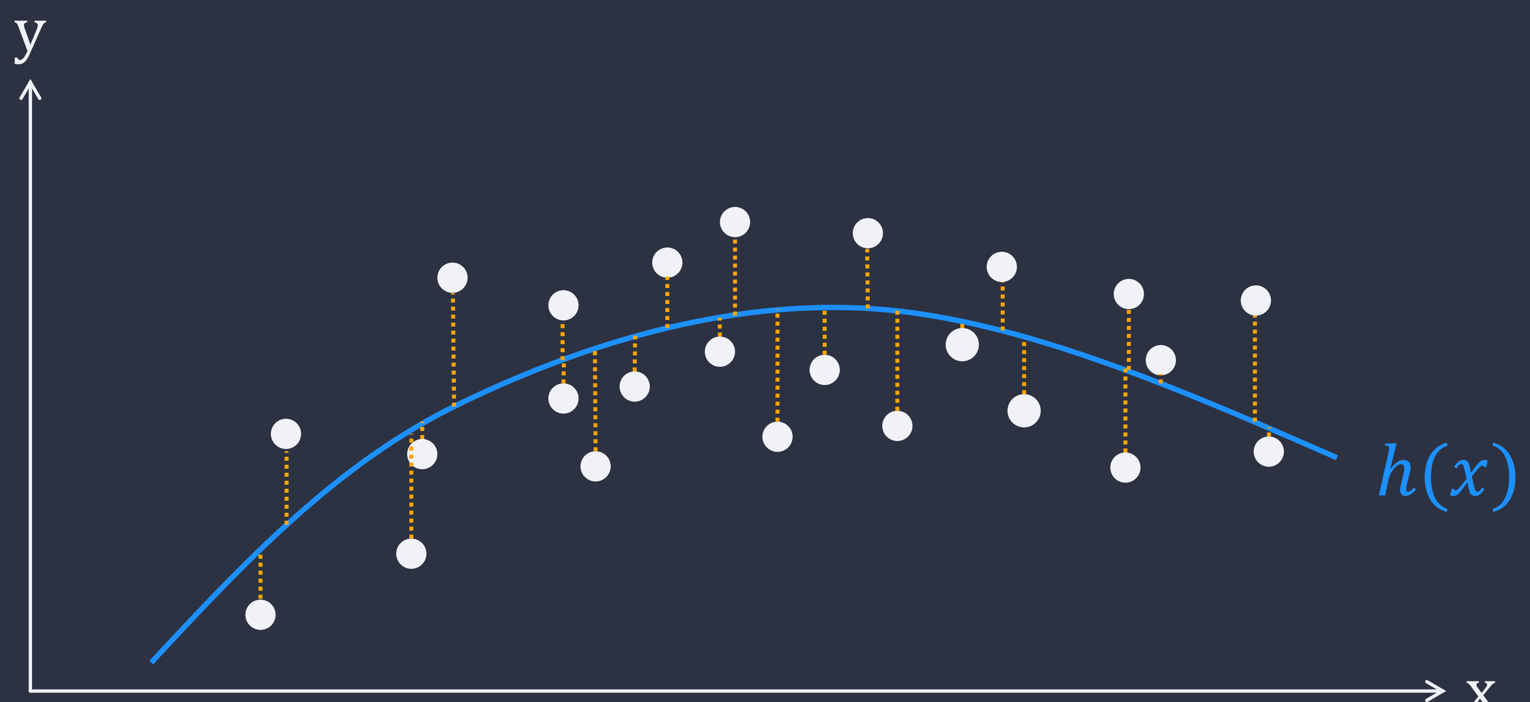
$$J = \frac{1}{m}\sum_{i=1}^{m}\left(y^{(i)} - \hat{y}^{(i)}\right)^{\boxed{2}}$$

Outlier dominates cost function



Mean Absolute Error (MAE)

Mean Absolute Error (MAE)

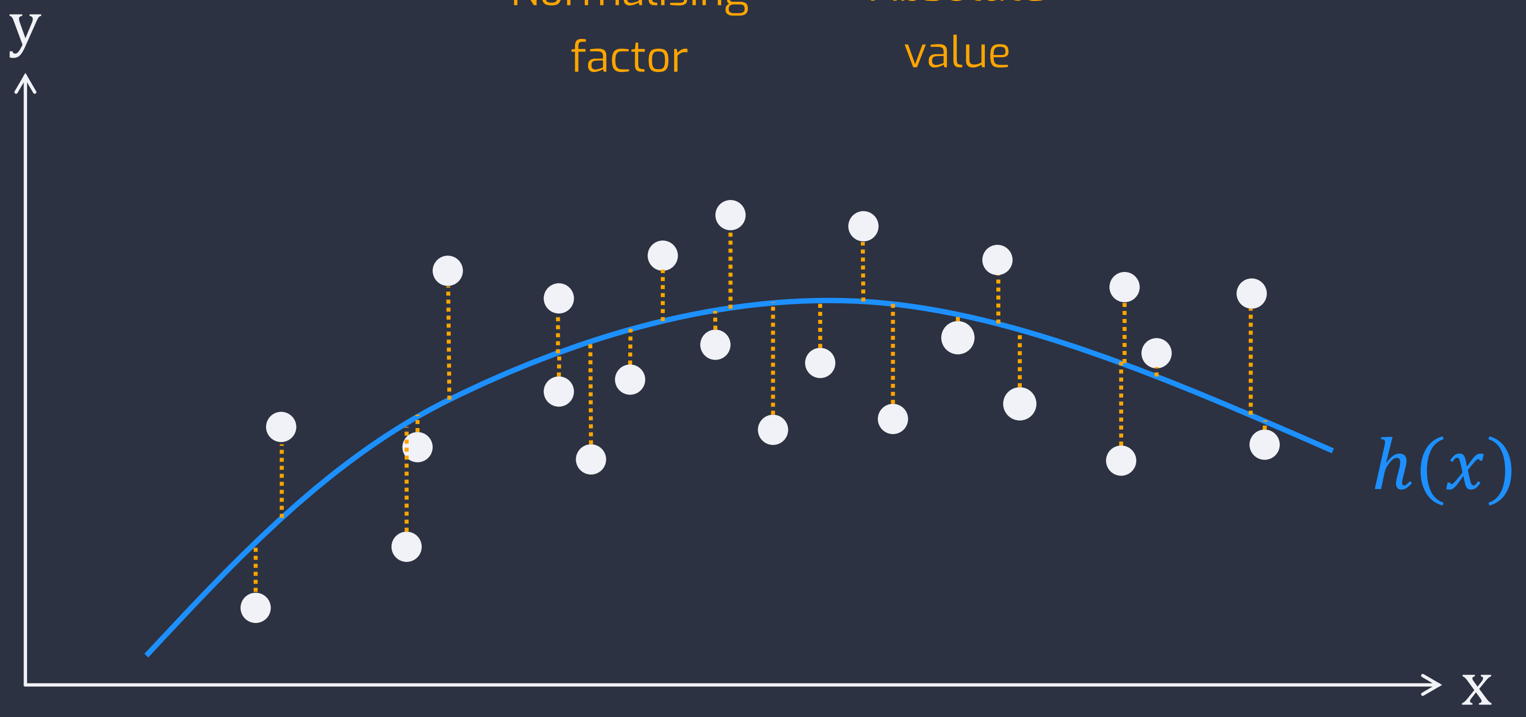


Mean Absolute Error (MAE)

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

Normalising factor

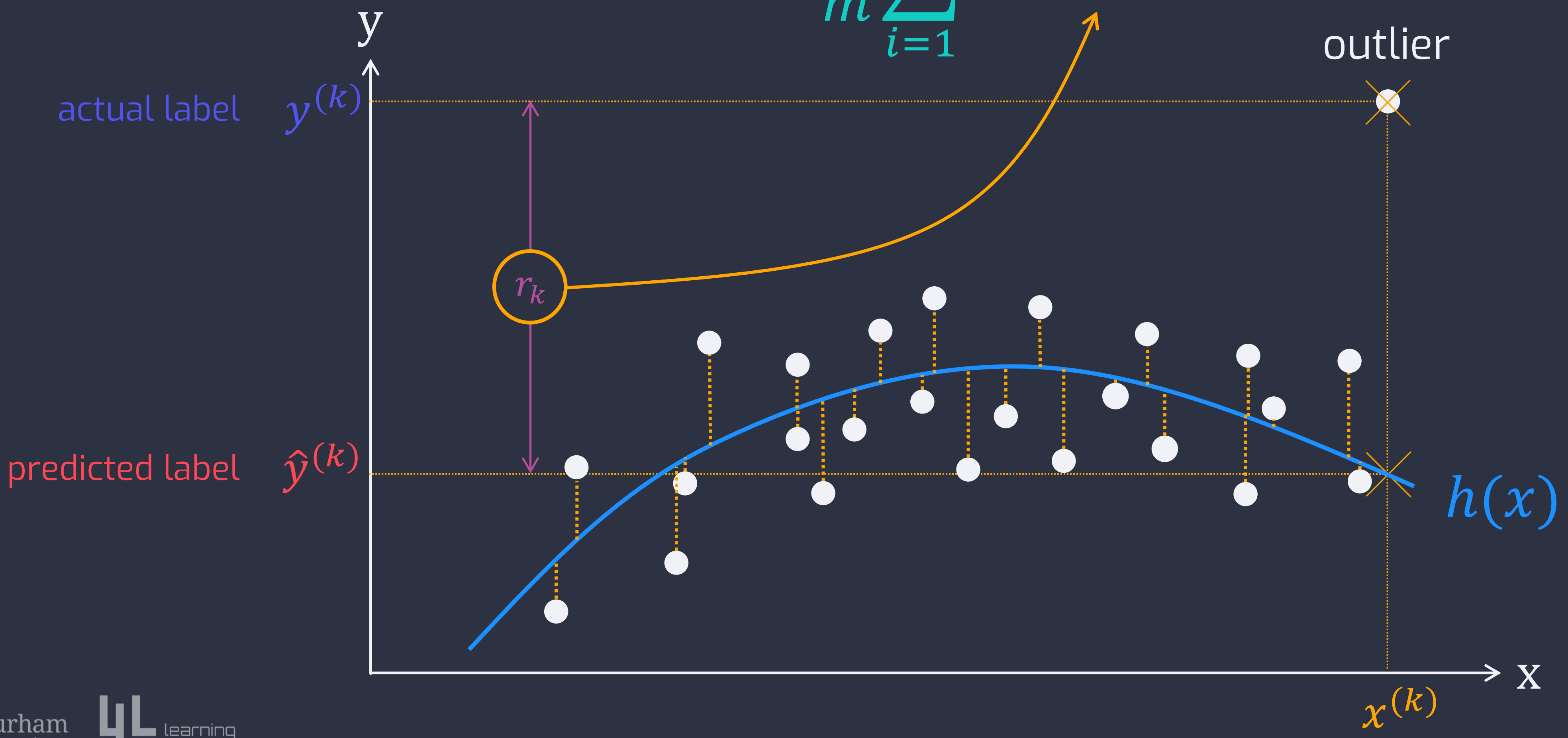
Absolute value



Mean Absolute Error (MAE)

- More robust with respect to outliers.

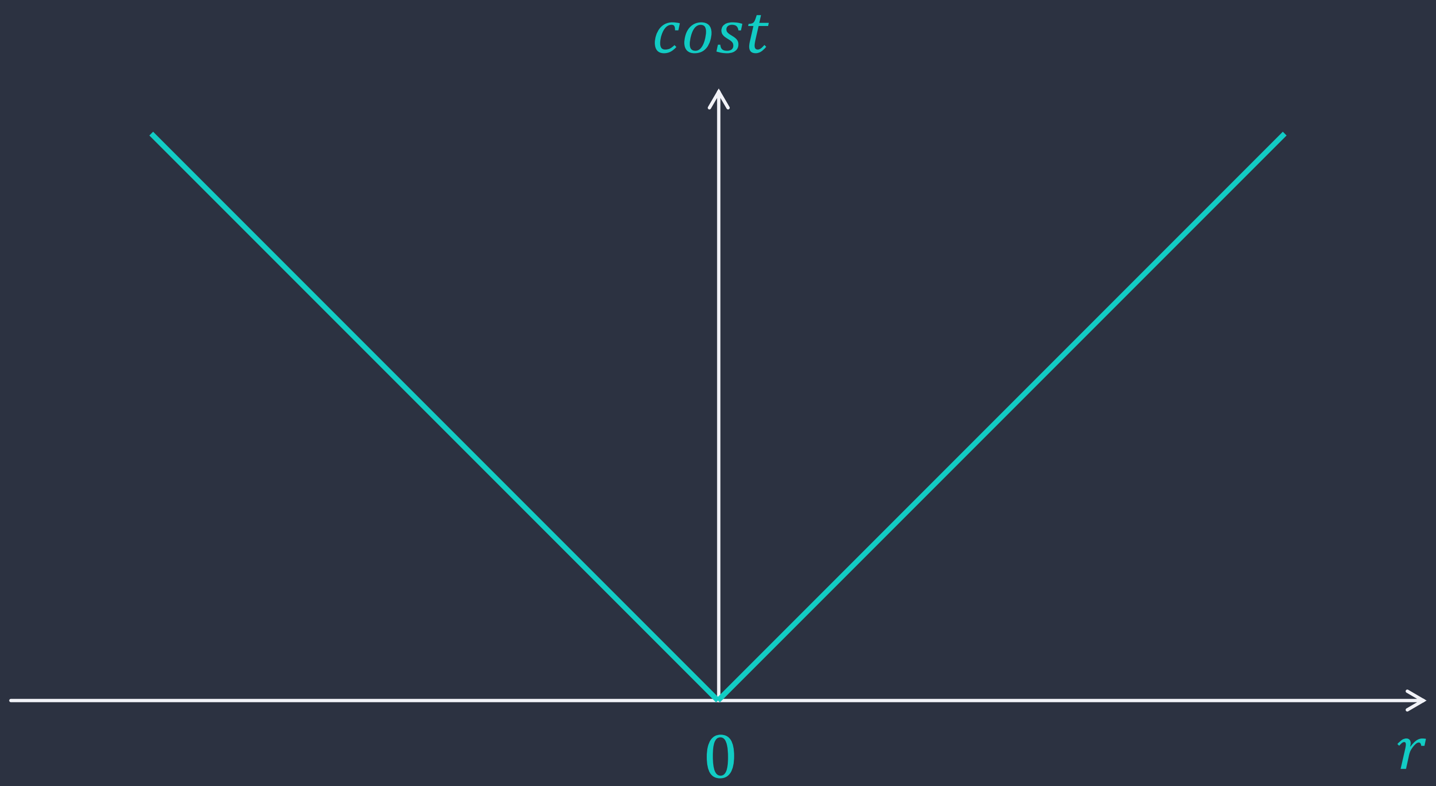
$$J = \frac{1}{m}\sum_{i=1}^{m} \left|y^{(i)} - \hat{y}^{(i)}\right|^{\cancel{2}}$$



Mean Absolute Error (MAE)

- More robust with respect to outliers.

$$J = \frac{1}{m}\sum_{i=1}^{m}|y^{(i)} - \hat{y}^{(i)}| \qquad \frac{d}{dr} = \begin{cases} -1, & r < 0 \\ +1, & r > 0 \end{cases} \qquad (r = y - \hat{y}, residual)$$

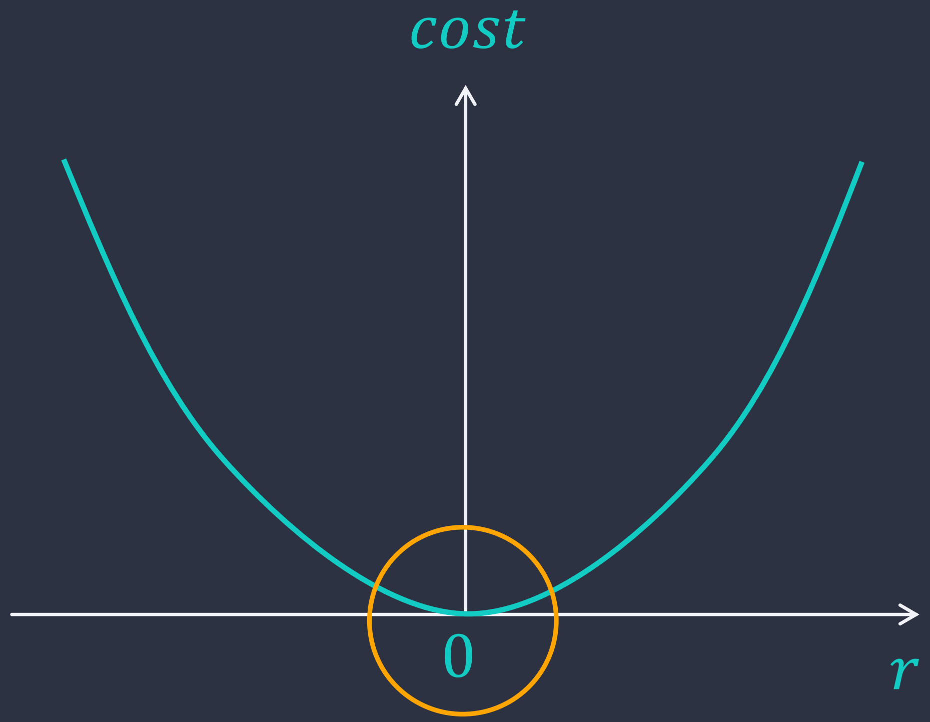


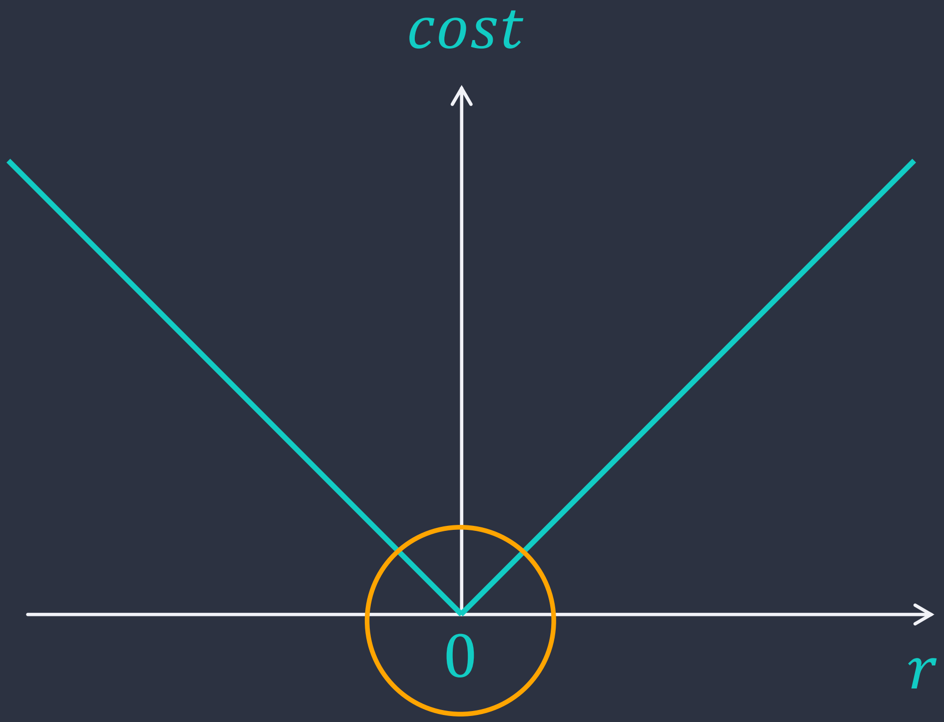
Mean Squared Error vs Mean Absolute Error

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

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

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





Reach minima when actual value (y) is exactly equal to predicted value $(\hat{y})$, i.e. $r = y - \hat{y} = 0$.

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

with slight variance

$index$	$error$	$error^2$	$\|error\|$
1	0	0	0
2	0.5	0.25	0.5
3	-1	1	1
4	1.5	2.25	1.5
5	-2	4	2

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

with slight variance

$$J_{MSE} = \frac{1}{5} \cdot (0 + 0.25 + 1 + 2.25 + 4) = 1.5$$

$$J_{MAE} = \frac{1}{5} \cdot (0 + 0.5 + 1 + 1.5 + 2) = 1$$

$index$		$error$	$error^2$	$\lvert error \rvert$
1		0	0	0
2		0.5	0.25	0.5
3		-1	1	1
4		1.5	2.25	1.5
5		-2	4	2
6	outlier	20	400	20

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

with slight variance

$$J_{MSE} = \frac{1}{5} \cdot (0 + 0.25 + 1 + 2.25 + 4) = 1.5$$

$$J_{MAE} = \frac{1}{5} \cdot (0 + 0.5 + 1 + 1.5 + 2) = 1$$

with outlier

$$J_{MSE} = \frac{1}{6} \cdot (0 + 0.25 + 1 + 2.25 + 4 + 400) = 67.92$$

$$J_{MAE} = \frac{1}{6} \cdot (0 + 0.5 + 1 + 1.5 + 2 + 20) = 4.17$$

$index$	$error$	$error^2$	$\lvert error \rvert$
1	0	0	0
2	0.5	0.25	0.5
3	-1	1	1
4	1.5	2.25	1.5
5	-2	4	2
6 outlier	20	400	20

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

with slight variance

$$J_{MSE} = \frac{1}{5} \cdot (0 + 0.25 + 1 + 2.25 + 4) = 1.5$$

$$J_{MAE} = \frac{1}{5} \cdot (0 + 0.5 + 1 + 1.5 + 2) = 1$$

with outlier

$$J_{MSE} = \frac{1}{6} \cdot (0 + 0.25 + 1 + 2.25 + 4 + 400) = 67.92$$

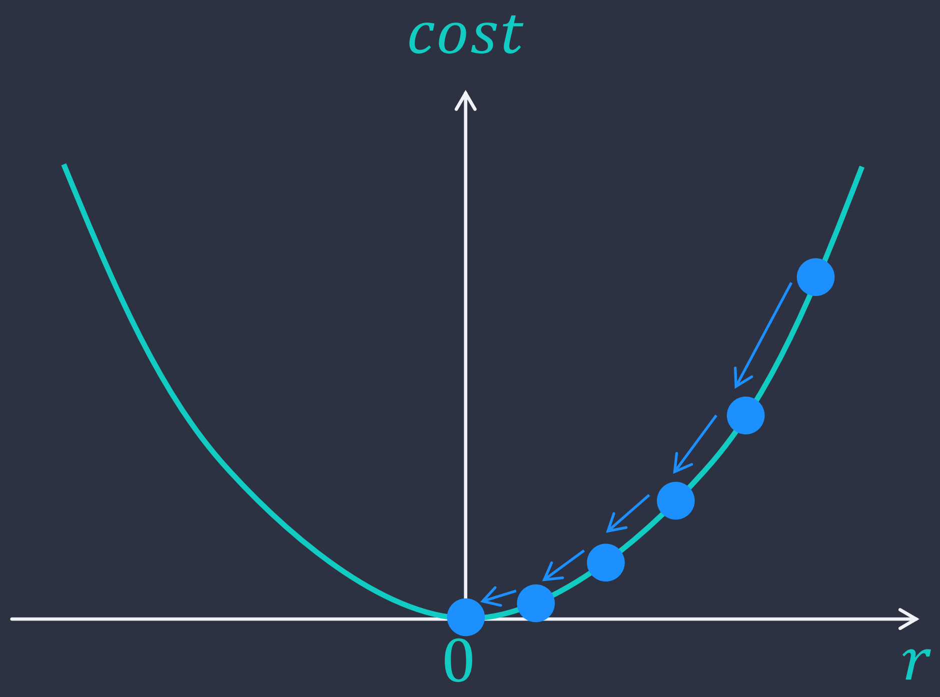
$$J_{MAE} = \frac{1}{6} \cdot (0 + 0.5 + 1 + 1.5 + 2 + 20) = 4.17$$

$$J_{RMSE} = J_{\sqrt{MSE}} = 8.24$$

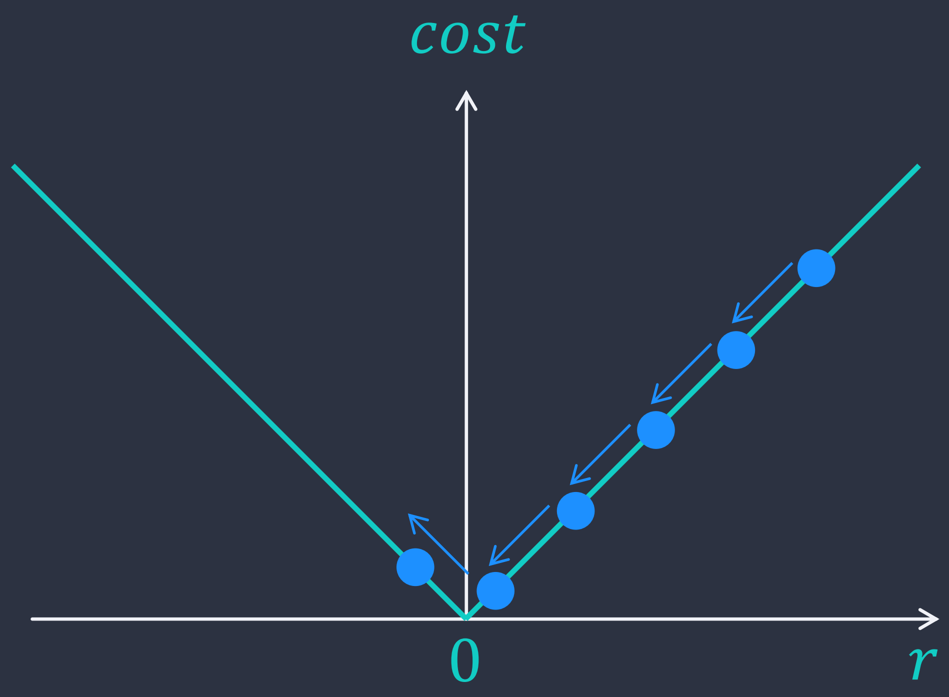
$index$		$error$	$error^2$	$\lvert error \rvert$
1		0	0	0
2		0.5	0.25	0.5
3		-1	1	1
4		1.5	2.25	1.5
5		-2	4	2
6	outlier	20	400	20

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

A big issue in MAE!



gradient becomes smaller
with fixed learning rate



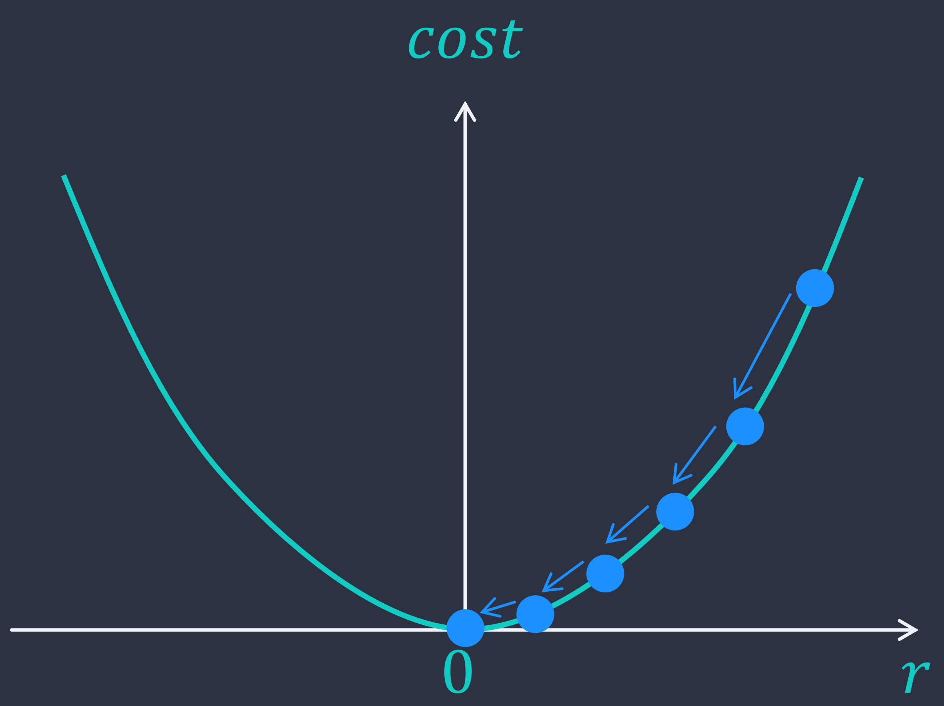
gradient remains the same
with fixed learning rate

Could make it dynamic – decrease as approaching 0.

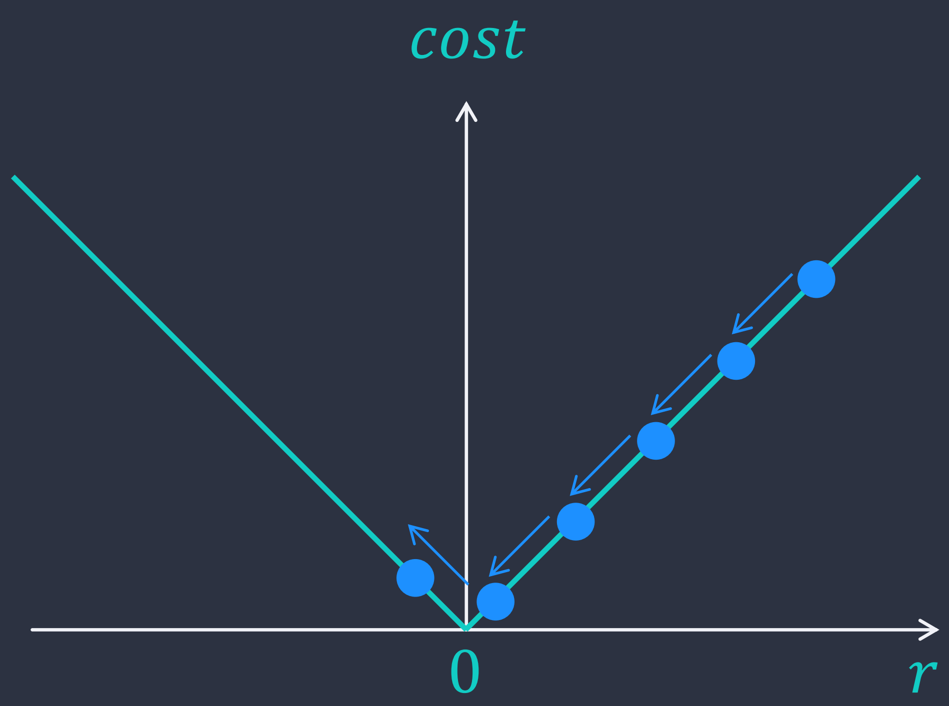
Mean Squared Error (MSE) vs Mean Absolute Error (MAE)

When to use which?

When outliers represent anomalies



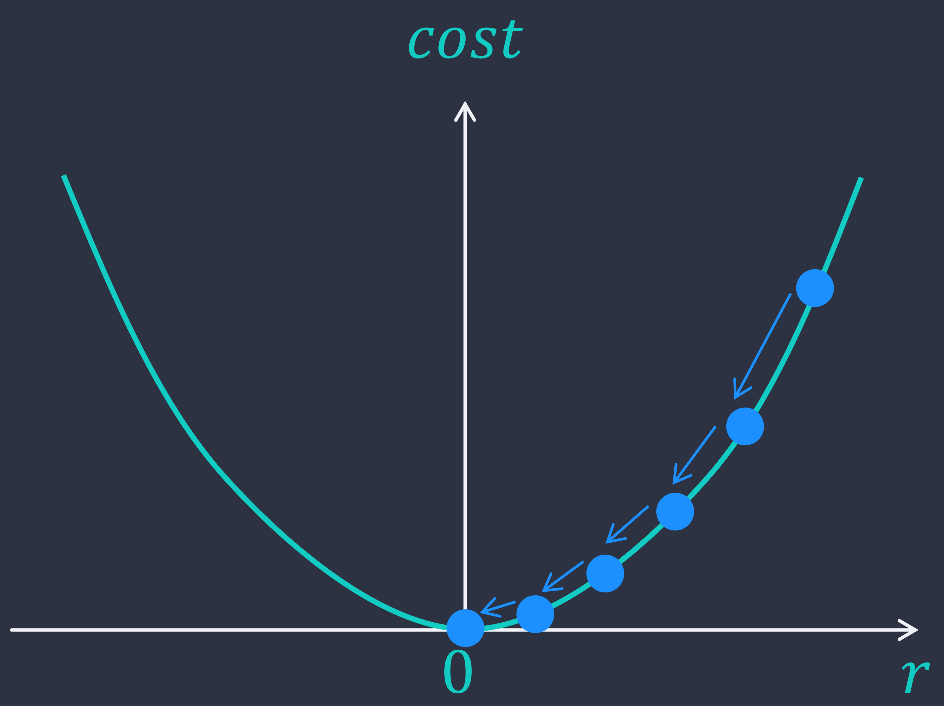
When outliers represent corrupted data



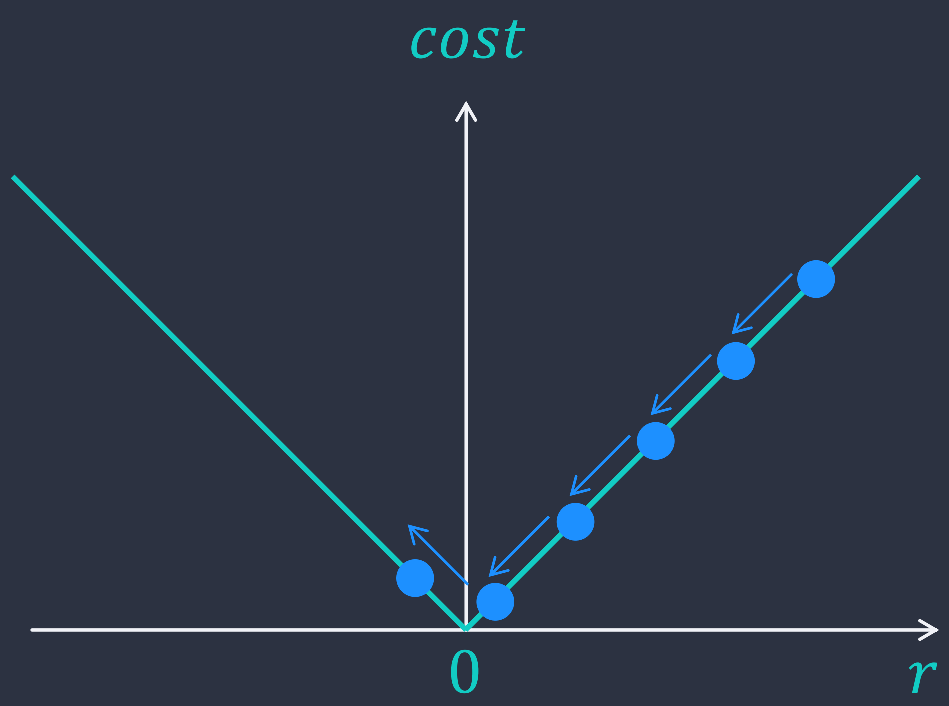
Mean Squared Error (MSE) vs Mean Absolute Error (MAE)

Issue for both, when learning from skewed/imbalanced data.

Ignoring outliers and achieving unrealistic high accuracy.



Got skewed towards outliers, achieving low accuracy,

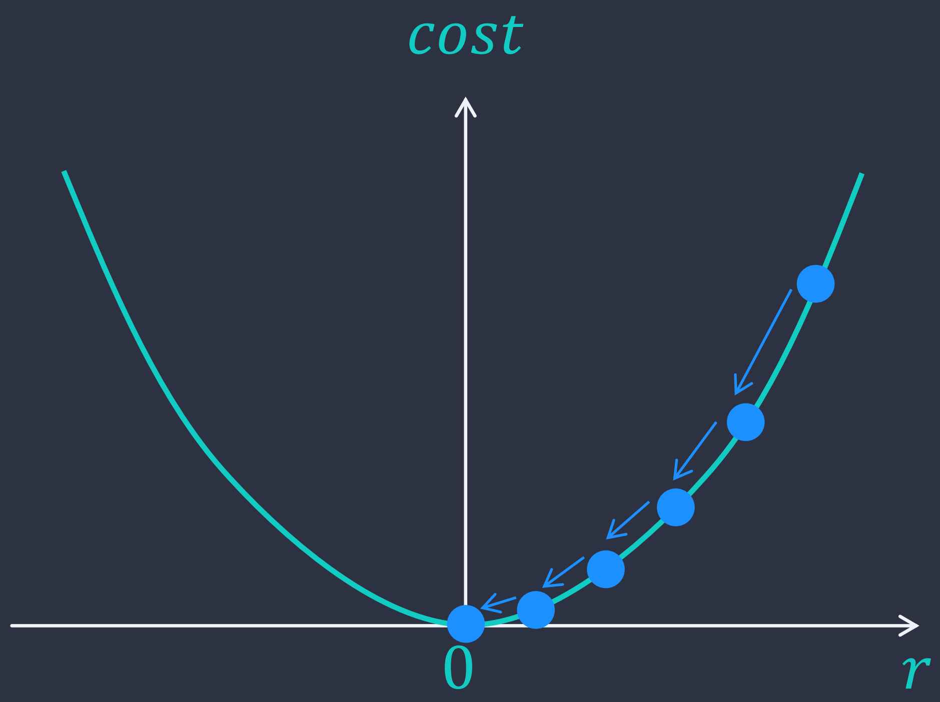


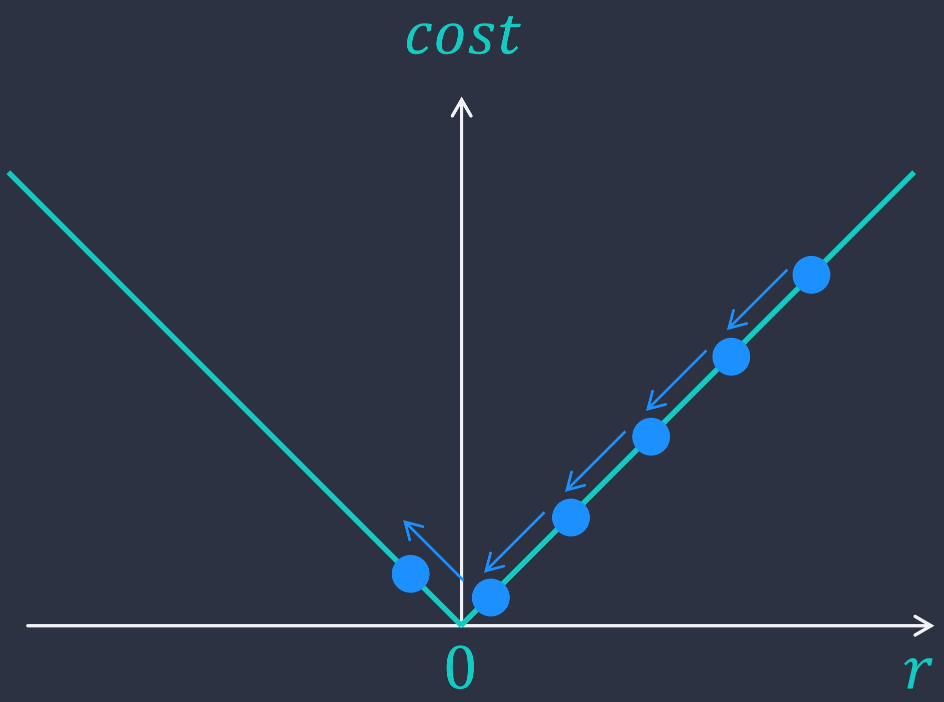
Mean Squared Error (MSE) vs Mean Absolute Error (MAE)

Issue for both, when learning from skewed/imbalanced data.

Solutions: data transformation,

or, Huber-M (in the next video)





Takeaway Points

- Cost function should be able to test model and make sure cost becomes smaller as model (hypothesis function) fits data better.
- MSE is intuitive and easy to implement but sensitive to outliers.
- MAE is more robust with respect to outliers but may pose computational challenge - not differentiable when error=0.
- With MAE, gradient remains the same – bad from learning.
- To use MSE if outliers represent anomalies; to use MAE if outliers represent corrupted data.
- Both MSE & MAE may have issues with skewed/imbalanced data.