COMP2261 ARTIFICIAL INTELLIGENCE / MACHINE LEARNING

Cost Functions

For Regression Models

-- Quantile

Dr SHI Lei

Durham University

learning lab

Learning Objectives

- Understand what is prediction interval
- Understand how Quantile cost function works
- Understand why use Quantile cost

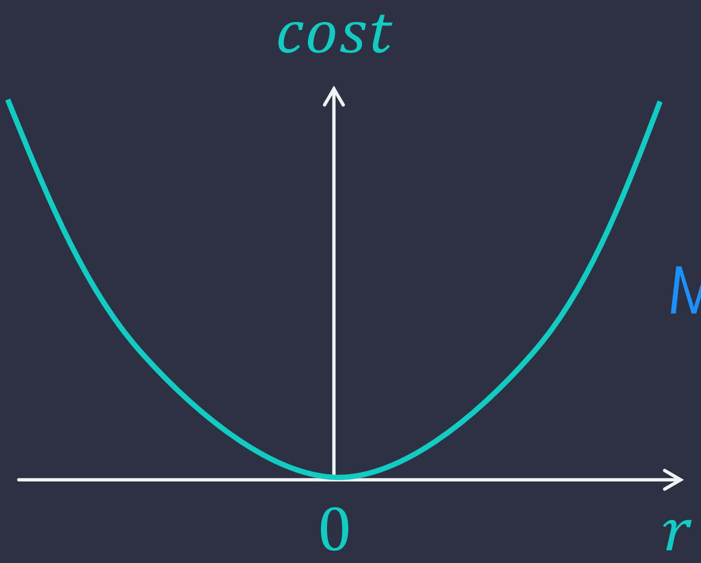
MSE

Quantile

MAE

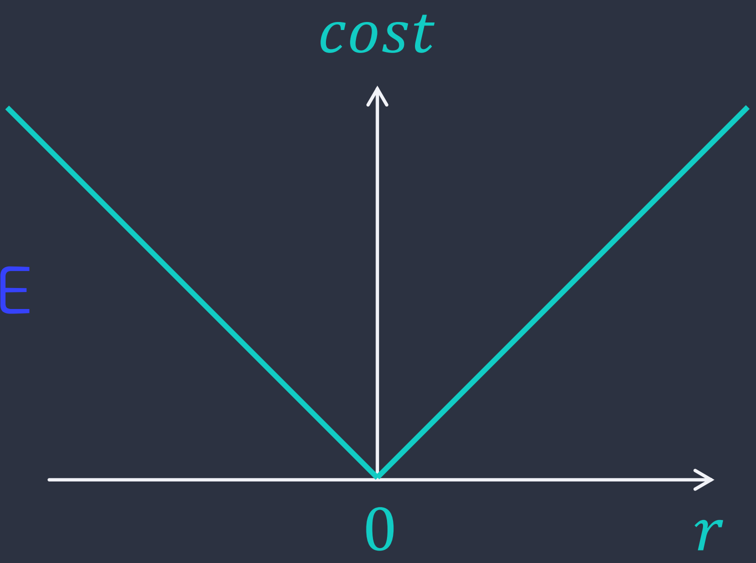
Log-Cosh

Huber-M

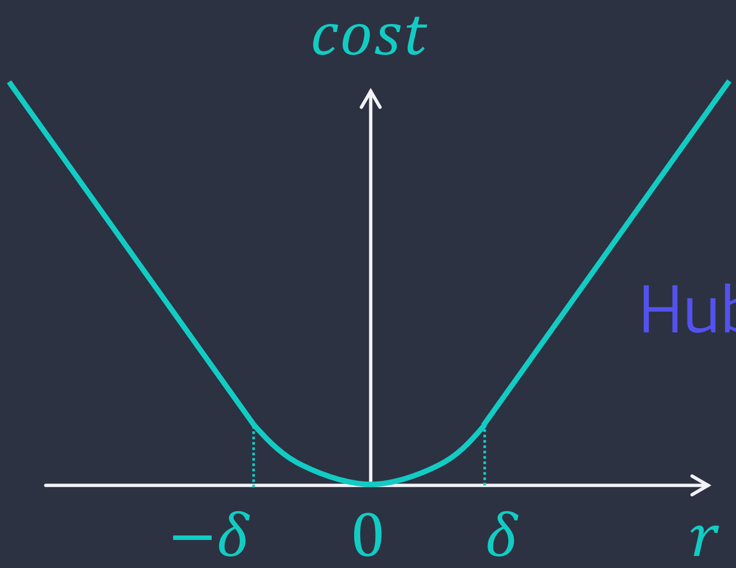


MSE

MAE

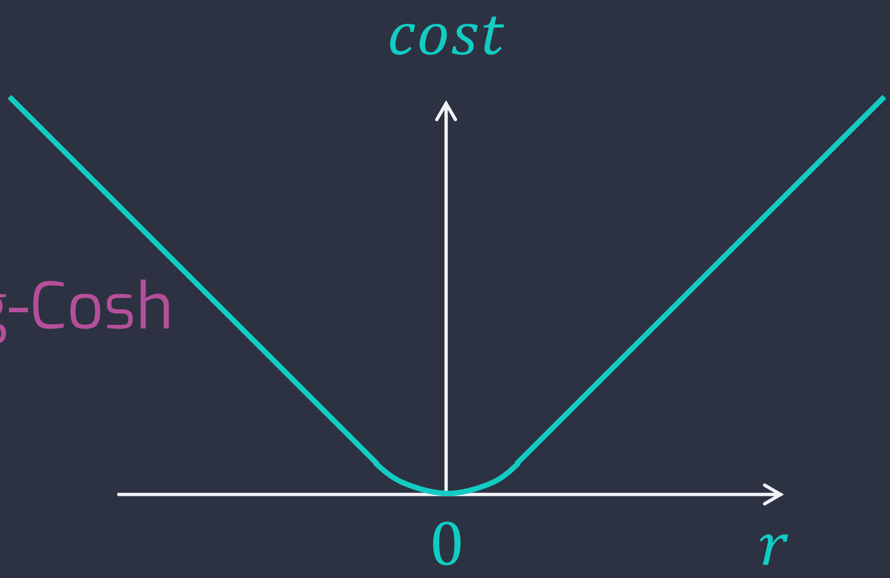


Point Prediction



Huber-M

Log-Cosh



Alternative to Point Prediction ?

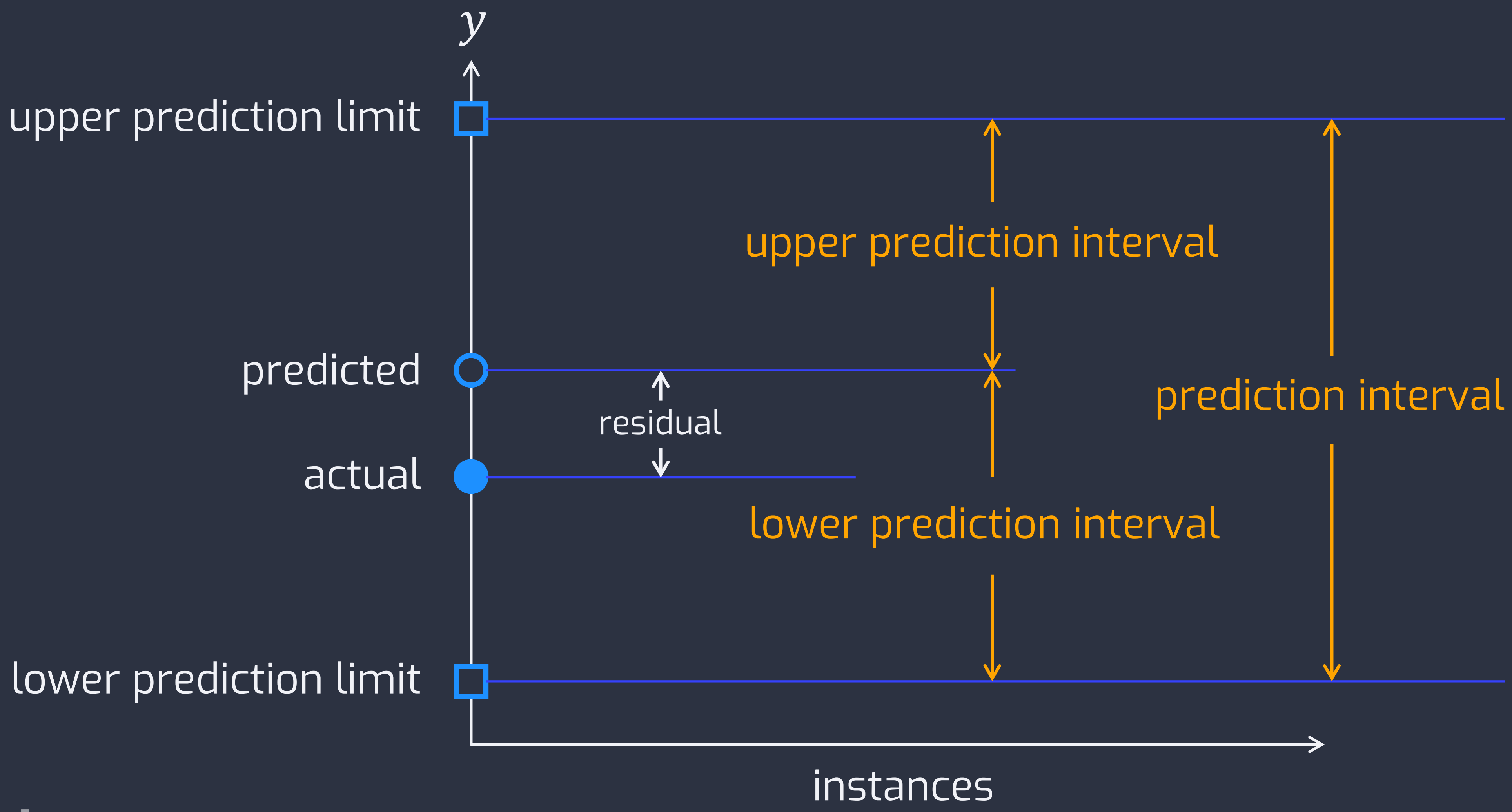
Prediction Interval

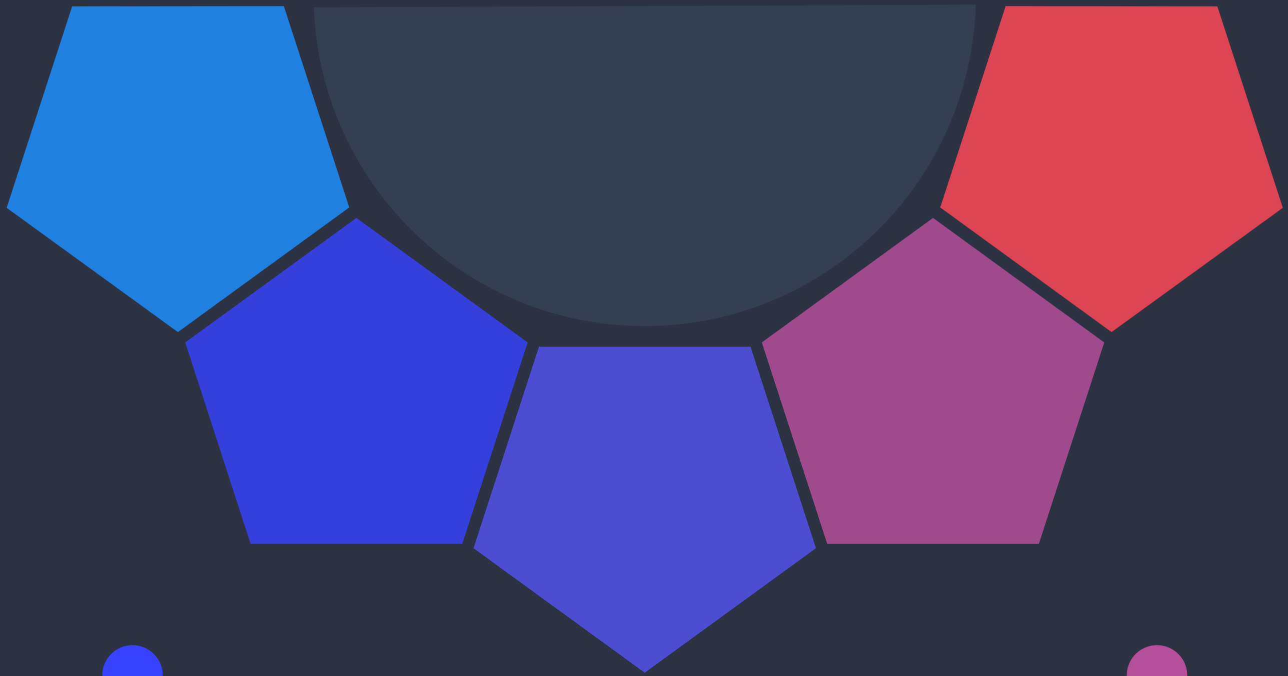
Alternative to Point Prediction

- Instead of prediction an single value, interval prediction predicts an interval in which a future instance will fall, with a certain probability.
- Different from confidence interval
  - Confidence Interval: quantifies uncertainty on an predicted population variable e.g. mean, standard deviation.
  - Prediction Interval: quantifies the uncertainty on a single instance predicted from the population.

Prediction Interval

-- a quantification of the uncertainty on a prediction





MSE

Quantile

MAE

Log-Cosh

Huber-M

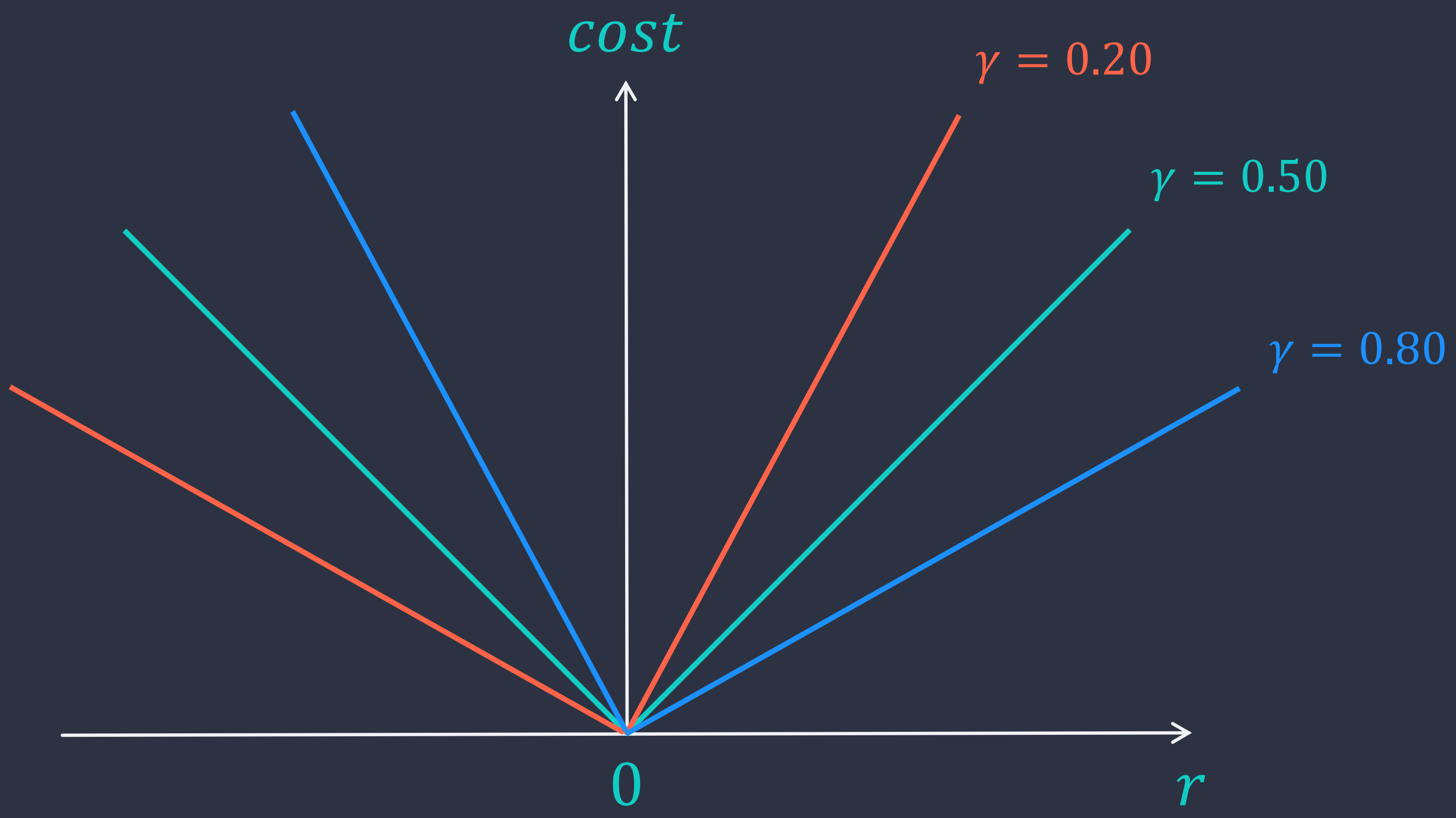
Quantile

Quantile Cost Function

$$J = \frac{1}{m}\Big(\sum_{y^{(i)}<\hat{y}^{(i)}} (\gamma - 1)\left|y^{(i)} - \hat{y}^{(i)}\right| + \sum_{y^{(i)}>\hat{y}^{(i)}} \gamma\left|y^{(i)} - \hat{y}^{(i)}\right|\Big) \qquad \gamma \in (0,1)$$

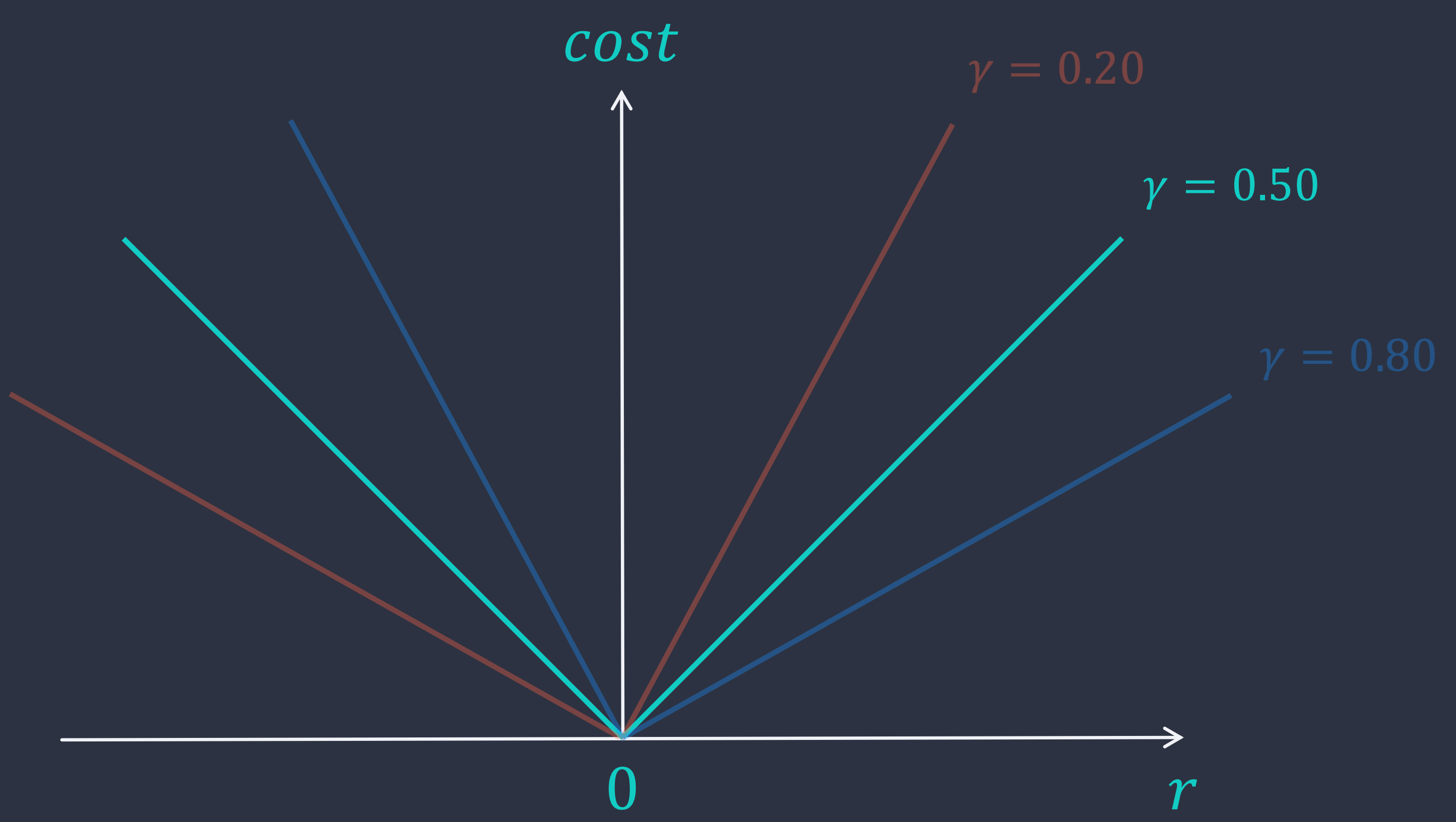
Quantile Cost Function

$$J = \frac{1}{m}\left(\sum_{y^{(i)}<\hat{y}^{(i)}} (\gamma - 1)\left|y^{(i)} - \hat{y}^{(i)}\right| + \sum_{y^{(i)}>\hat{y}^{(i)}} \gamma\left|y^{(i)} - \hat{y}^{(i)}\right|\right) \qquad \gamma \in (0,1)$$



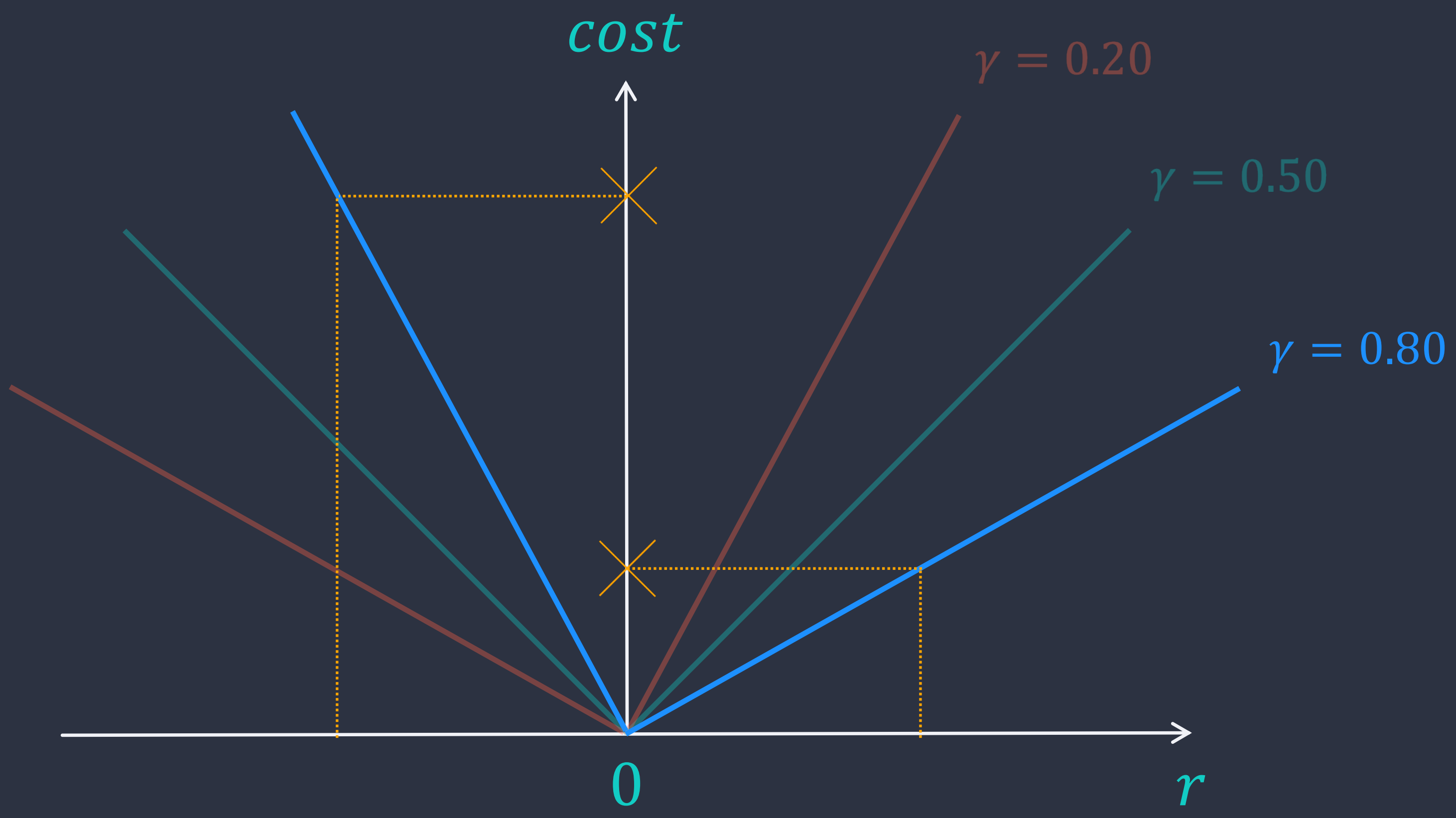
Quantile Cost Function

$$J = \frac{1}{m}\Big(\sum_{y^{(i)}<\hat{y}^{(i)}} (\gamma - 1)\left|y^{(i)} - \hat{y}^{(i)}\right| + \sum_{y^{(i)}>\hat{y}^{(i)}} \gamma\left|y^{(i)} - \hat{y}^{(i)}\right|\Big) \qquad \gamma \in (0,1)$$



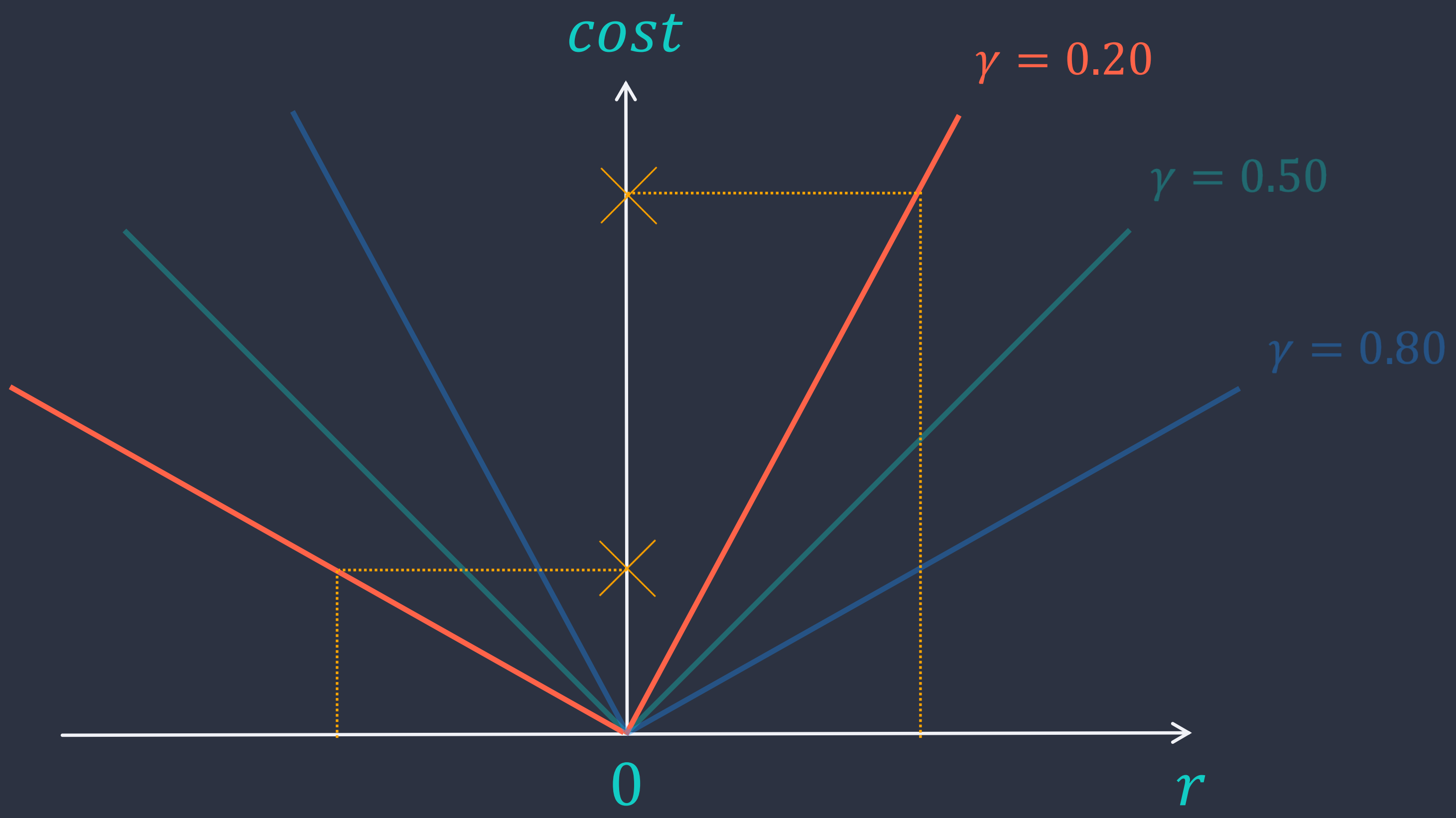
Quantile Cost Function

$$J = \frac{1}{m}\left(\sum_{y^{(i)} < \hat{y}^{(i)}} (\gamma - 1)\left|y^{(i)} - \hat{y}^{(i)}\right| + \sum_{y^{(i)} > \hat{y}^{(i)}} \gamma\left|y^{(i)} - \hat{y}^{(i)}\right| \right) \qquad \gamma \in (0,1)$$

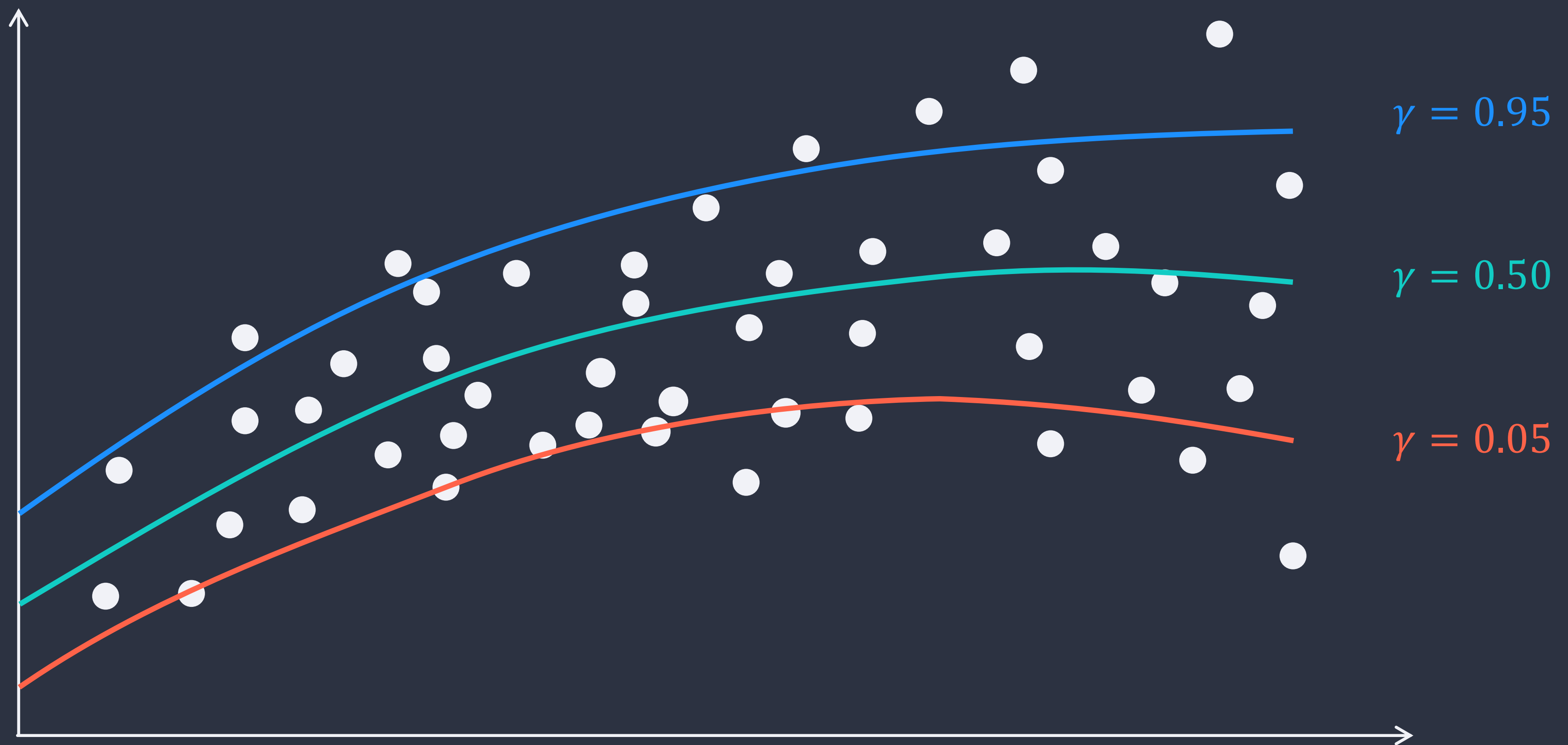


Quantile Cost Function

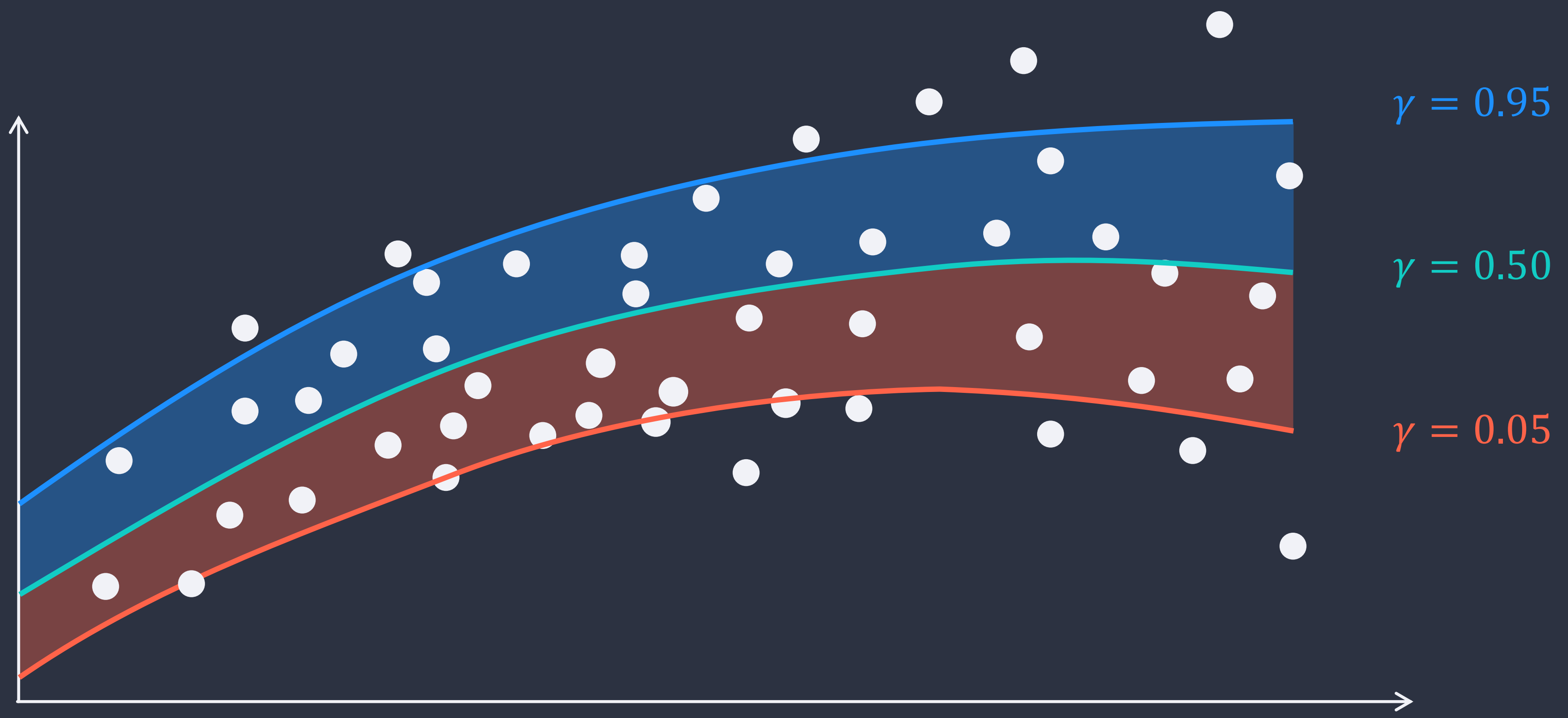
$$J = \frac{1}{m}\left(\sum_{y^{(i)}<\hat{y}^{(i)}} (\gamma - 1)\left|y^{(i)} - \hat{y}^{(i)}\right| + \sum_{y^{(i)}>\hat{y}^{(i)}} \gamma\left|y^{(i)} - \hat{y}^{(i)}\right| \right) \qquad \gamma \in (0,1)$$



In Practice



In Practice



Quantile regression for the 5th and 95th quantiles tries to find bounds $y_0(\boldsymbol{x})$ and $y_1(\boldsymbol{x})$, on the dependent variable y given predictor variable $\boldsymbol{x}$, such that

$$\mathbb{P}\big(Y \le y_0(\boldsymbol{X})\big) = 0.05 \qquad \mathbb{P}\big(Y \le y_1(\boldsymbol{X})\big) = 0.95$$

So, $\mathbb{P}\big(y_0(\boldsymbol{X})Y \le y_1(\boldsymbol{X})\big) = 0.90$ (a 90% prediction interval)

Takeaway Points

- Point predictions are based on an assumption that residuals have constant variance across values of independent variables.
- Output of regression models are subject to uncertainty which can be modelled by prediction intervals.
- Prediction interval can measure the likeliness of correctness of predictions, and thus providing probabilistic prediction limits.
- Quantile cost function can be useful when we are interested in predicting an interval instead of only point predictions.
- Quantile cost function can be used for data with residuals that have non-constant variance or non-normal distribution.