Regression metrics:
(R)MSE, R-squared, MAE

Plan for the video

1) **Regression**
   - MSE, RMSE, R-squared
   - MAE
   - (R)MSPE, MAPE
   - (R)MSLE
2) **Classification:**
   - Accuracy, LogLoss, AUC
   - Cohen's (Quadratic weighted) Kappa

Plan for the video

1) **Regression**
   - MSE, RMSE, R-squared
   - MAE
   - (R)MSPE, MAPE
   - (R)MSLE
2) **Classification:**
   - Accuracy, LogLoss, AUC
   - Cohen's (Quadratic weighted) Kappa

Notation

- N – number of objects

- $y \in \mathbb{R}^N$ – target values

  $\hat{y} \in \mathbb{R}^N$ – predictions

- $\hat{y}_i \in \mathbb{R}$ – prediction for i-th object

  $y_i \in \mathbb{R}$ – target for i-th object

MSE: Mean Square Error

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

MSE: Mean Square Error

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

Data:

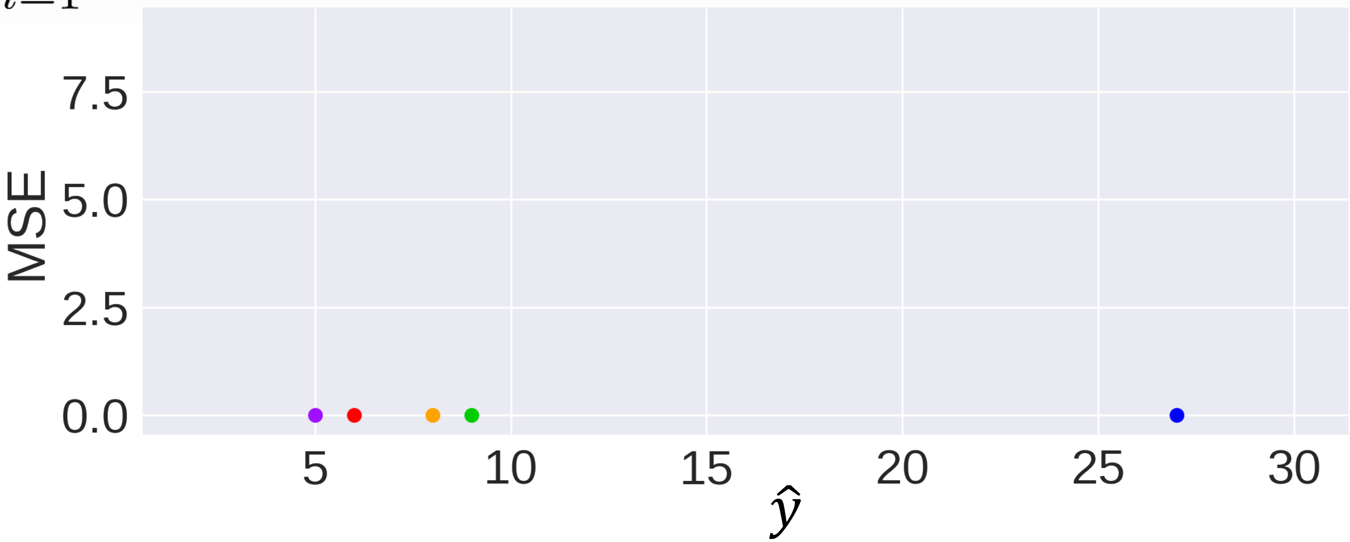
X	Y
...	5
...	9
...	8
...	6
...	27

MSE: Mean Square Error

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

Data:

X	Y
...	5
...	9
...	8
...	6
...	27

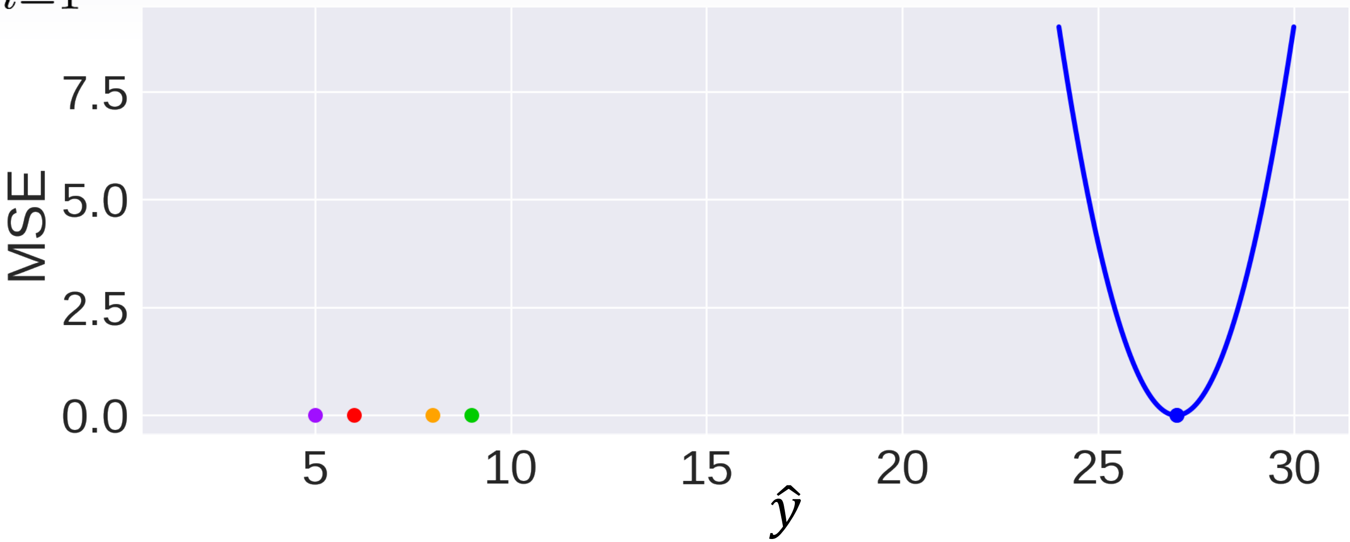


MSE: Mean Square Error

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

Data:

X	Y
...	5
...	9
...	8
...	6
...	27

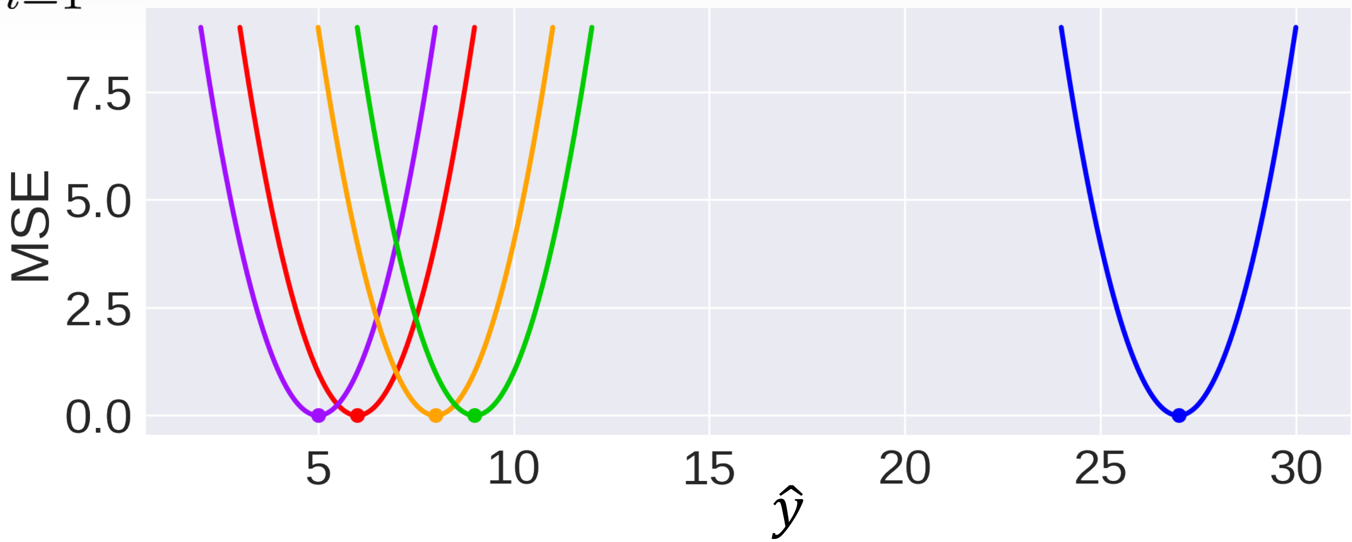


MSE: Mean Square Error

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

Data:

X	Y
...	5
...	9
...	8
...	6
...	27



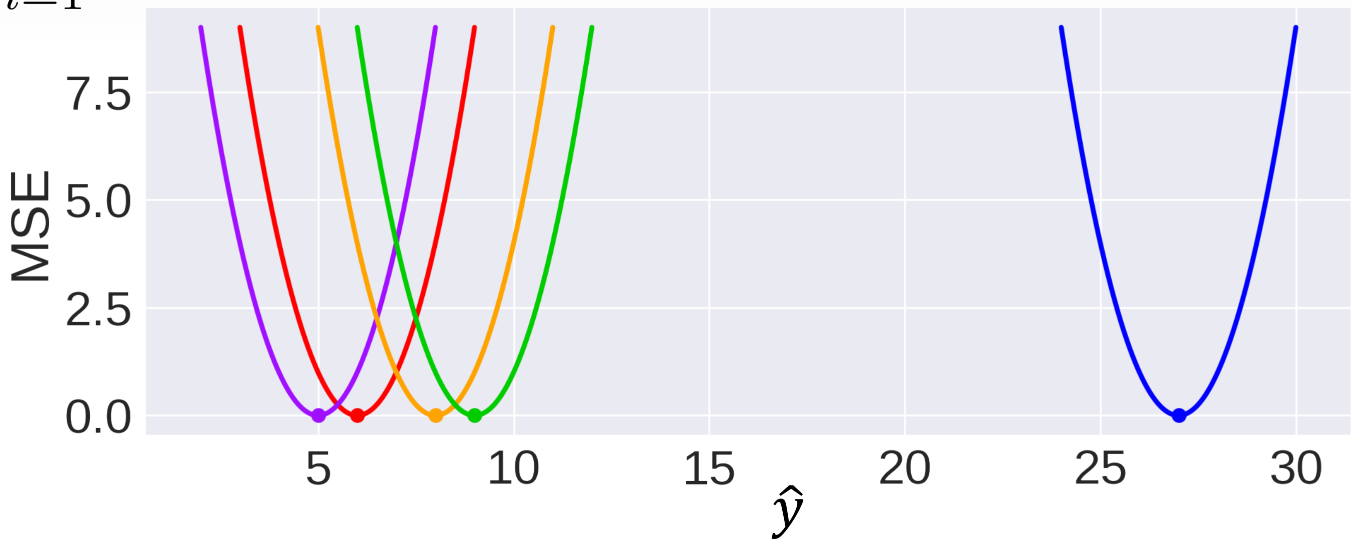
MSE: optimal constant

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \alpha)^2$$

Best constant: ?

Data:

X	Y
...	5
...	9
...	8
...	6
...	27



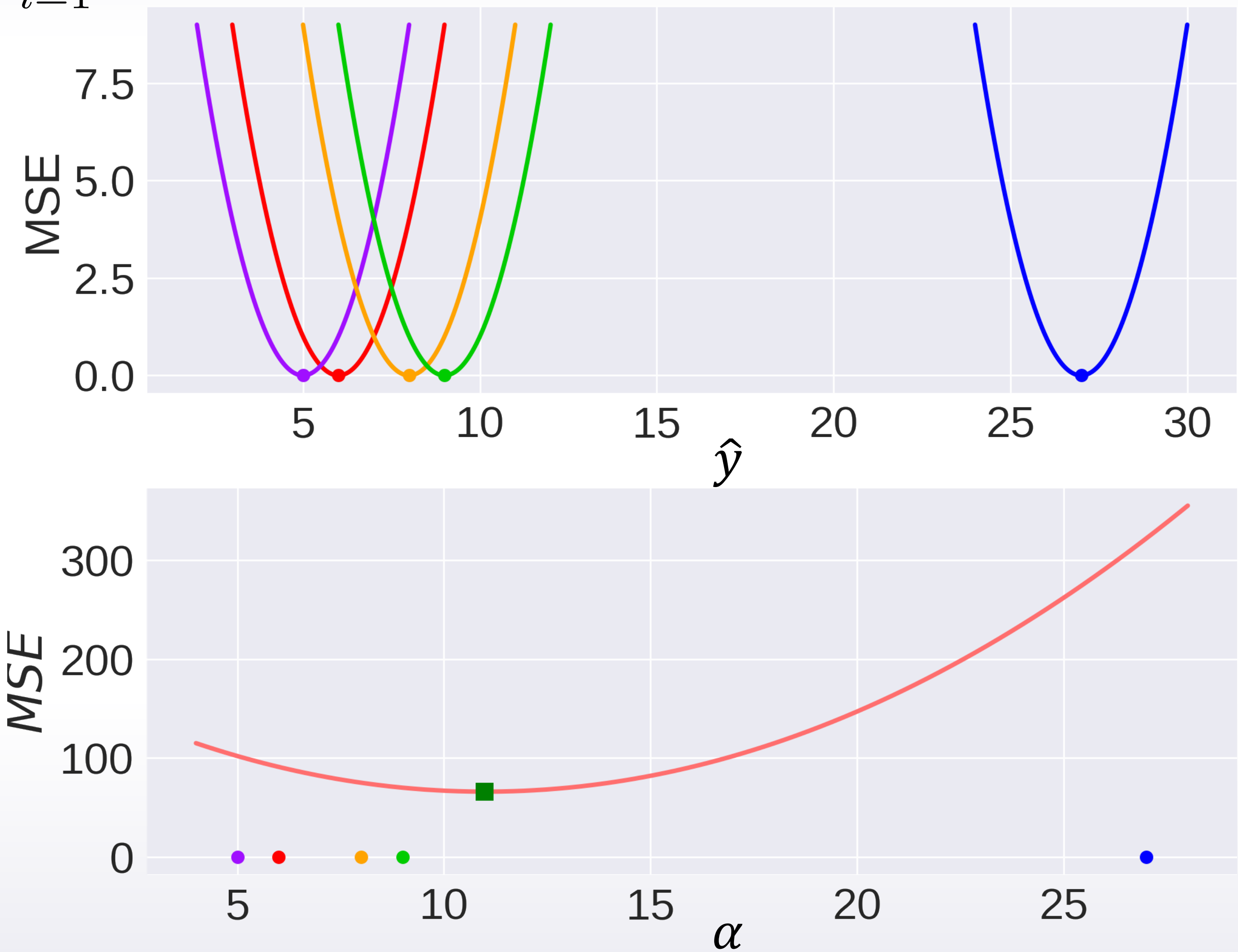
MSE: optimal constant

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \alpha)^2$$

Best constant: target mean

Data:

X	Y
...	5
...	9
...	8
...	6
...	27



MSE notes: RMSE

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

MSE notes: RMSE

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

Root mean square error

- $$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} = \sqrt{\text{MSE}}$$

MSE notes: RMSE

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

Root mean square error

- $$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} = \sqrt{\text{MSE}}$$
- $$\text{MSE}(a) > \text{MSE}(b) \Longleftrightarrow \text{RMSE}(a) > \text{RMSE}(b)$$

MSE notes: RMSE

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

Root mean square error

- $$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} = \sqrt{\text{MSE}}$$
- $$\text{MSE}(a) > \text{MSE}(b) \iff \text{RMSE}(a) > \text{RMSE}(b)$$
- $$\frac{\partial \text{RMSE}}{\partial \hat{y}_i} = \frac{1}{2\sqrt{MSE}}\frac{\partial \text{MSE}}{\partial \hat{y}_i}$$

MSE notes: RMSE

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

MSE notes: RMSE

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

R-squared:

$$R^2 = 1 - \frac{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2} = 1 - \frac{MSE}{\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

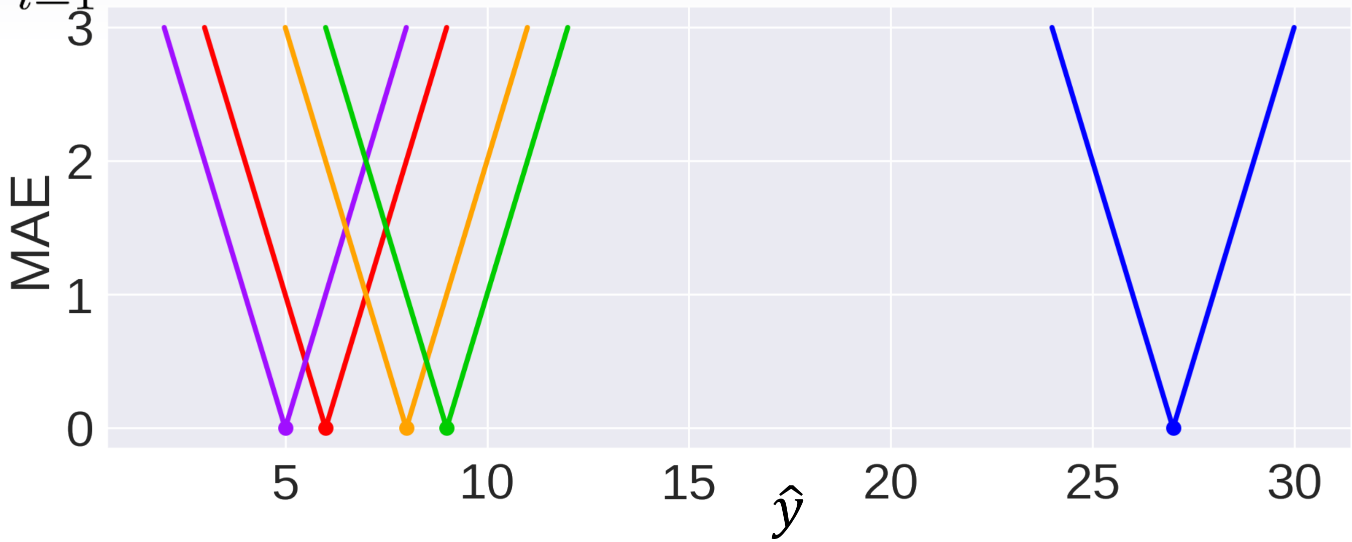
$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$$

MAE: Mean Absolute Error

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i|$$

Data:

X	Y
-1	5
1	9
-2	8
3	6
3	27



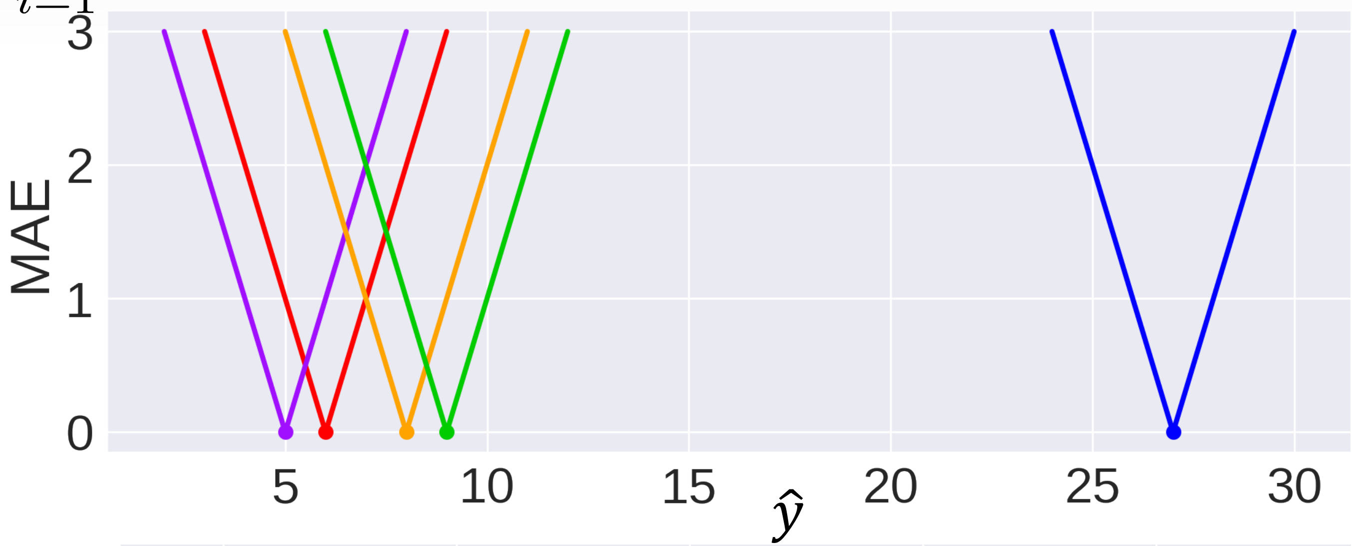
MAE: optimal constant

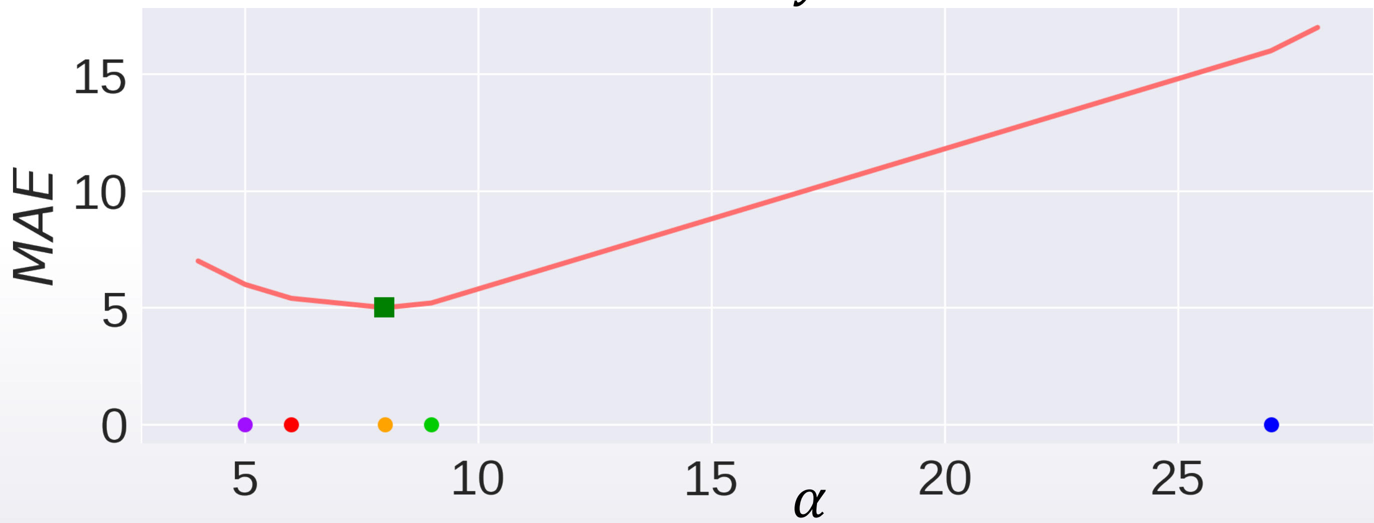
$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} |y_i - \alpha|$$

Best constant: target median

Data:

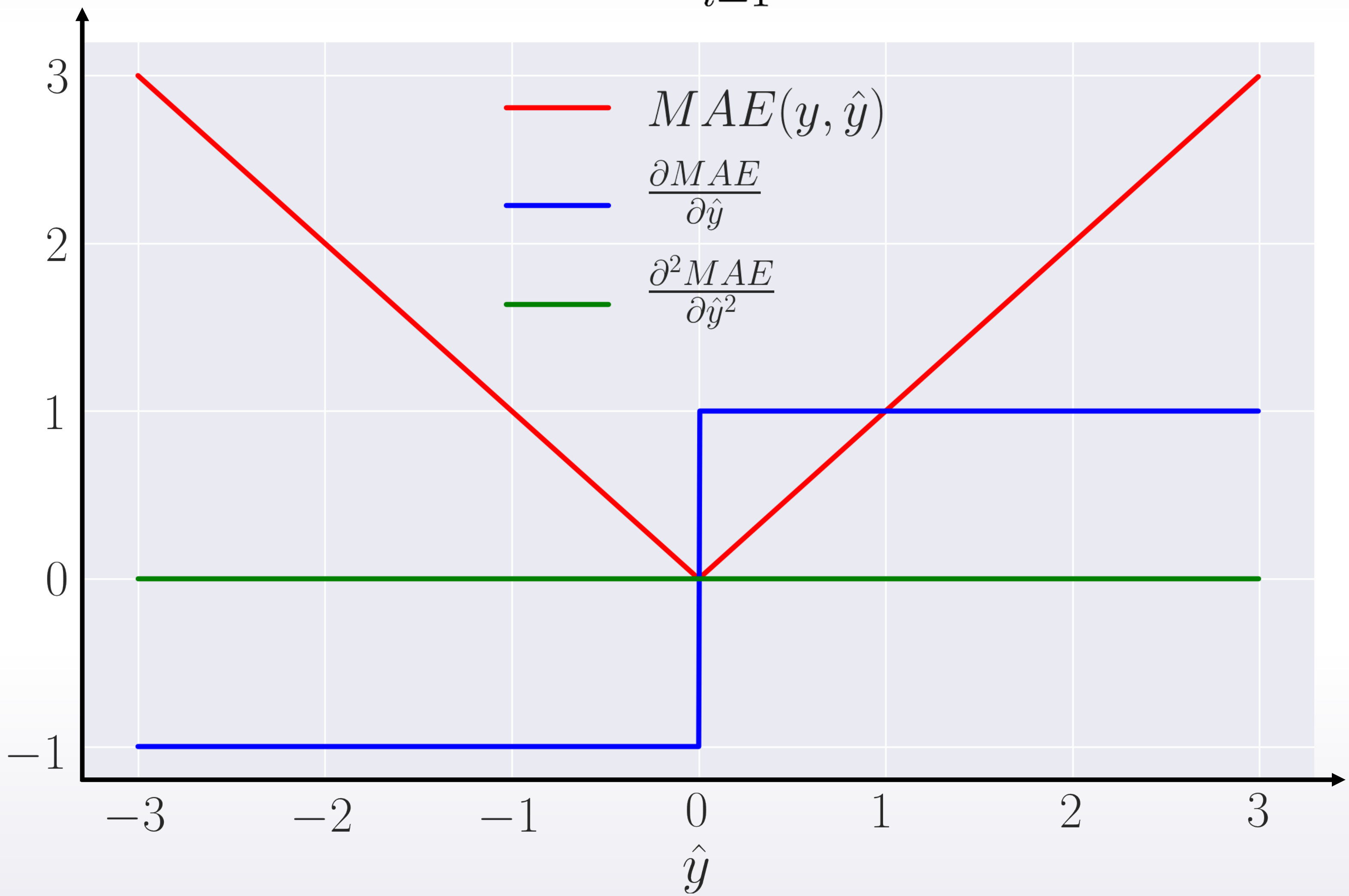
X	Y
-1	5
1	9
-2	8
3	6
3	27





MAE: derivatives

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|$$



MAE vs MSE

- **Do you have outliers in the data?**
  - Use MAE
- **Are you sure they are outliers?**
  - Use MAE
- **Or they are just unexpected values we should still care about?**
  - Use MSE

Conclusion

- Discussed the following metrics:
  - **MSE, RMSE, R-squared**
    - They are the same from optimization perspective
  - **MAE**
    - Robust to outliers