ERROR

ANALYSIS

Error Metrics

- actual $A = \left(a_1, \ldots . a_n\right)$
- predicted $P = \left(p_1, \ldots, p_n\right)$
- *max absolute error*

$$\max\left(|a_1 - p_1|, \ldots, |a_n - p_n|\right)$$

- *median absolute error*

$$\text{median}\left(|a_1 - p_1|, \ldots, |a_n - p_n|\right)$$

Error Metrics (cont'd)

- *mean absolute error* (MAE)

$$\frac{1}{n}\sum_{i=1}^{n}|a_i - p_i|$$

- *root mean squared error* (RMSE)

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(a_i - p_i)^2}$$

Error Computation

i	1	2	3	4	5
a_i	6	8	7	8	6
p_i	4	7	4	4	3
$\|a_i - p_i\|$	2	1	3	4	3
$(a_i - p_i)^2$	4	1	9	16	9

- max absolute error: 4
- median absolute error: 3

Error Computation (cont'd)

- mean absolute error

$$\frac{1}{n}\sum_{i=1}^{n}|a_i - p_i| = \frac{13}{5} = 2.6$$

- root mean squared error

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(a_i - p_i)^2} = \sqrt{\frac{39}{5}} \approx 2.8$$

Evenly Distributed Errors

i	1	2	3	4	5
a_i	6	8	7	8	6
p_i	5	9	6	7	5
$\|a_i - p_i\|$	1	1	1	1	1
$(a_i - p_i)^2$	1	1	1	1	1

$$\text{MAE} = \text{RMSE} = 1$$

Small Error Variance

i	1	2	3	4	5
a_i	6	8	7	8	6
p_i	5	9	6	6	4
$\|a_i - p_i\|$	1	1	1	2	2
$(a_i - p_i)^2$	1	1	1	4	4

$$\mathrm{MAE} = \frac{7}{5} = 1.40$$

$$\mathrm{RMSE} = \sqrt{\frac{11}{5}} \approx 1.48$$

Large Error Outlier

i	1	2	3	4	5
a_i	6	8	7	8	6
p_i	5	9	27	6	4
$\lvert a_i - p_i \rvert$	1	1	20	2	2
$(a_i - p_i)^2$	1	1	400	4	4

$$\text{MAE} = \frac{26}{5} = 5.2$$

$$\text{RMSE} = \sqrt{\frac{410}{5}} \approx 9.06$$

MAE vs. RMSE

- MAE $\leq$ RMSE
- RMSE $\leq$ MAE $\cdot \sqrt{n}$
- (RMSE - MAE)$\uparrow$ as $n \uparrow$
- MAE: easy to interpret
- RMSE: no absolute values