Demonstrating that the Mean Absolute Error is bounded by the Root Mean Square Error in prediction.

Benjamin Ciotti, 12/14/2024, BenCiotti.Stats@gmail.com

Suppose that for a sample of size n we have y_i being the observed values and $\widehat{y_i}$ being the predicted values of our dependent variable. Here $i = 1, 2, \ldots, n$.

Then the mean absolute error

$$MAE = \frac{1}{n}\sum |y_i - \widehat{y_i}|$$

and the root mean square error

$$RMSE = \sqrt{\frac{1}{n}\sum |y_i - \widehat{y_i}|^2}\,.$$

The Cauchy-Schwarz Inequality (CS) states that

$$\sum |a_i b_i| \le \sqrt{\sum a_i^2} \cdot \sqrt{\sum b_i^2}$$

Letting $a_i = |y_i - \widehat{y_i}|$ and $b_i = 1$ we get that

$$\sum |a_i b_i| = \sum |y_i - \widehat{y_i}| \cdot 1 = n \cdot MAE,$$

$$\sqrt{\sum a_i^2} = \sqrt{\sum |y_i - \widehat{y_i}|^2} = \sqrt{n} \cdot RMSE,$$

$$\sqrt{\sum b_i^2} = \sqrt{\sum 1^2} = \sqrt{n}.$$

Hence CS gives

$$n \cdot MAE \le \sqrt{n} \cdot RMSE \cdot \sqrt{n}$$

i.e.

$$MAE \le RMSE.$$

Thus the claim is proven.

Note:

I believe CS is sharp exactly when there exists a number λ such that $b_i = \lambda a_i$ for all i.

In our case this means $1 = \lambda|y_i - \widehat{y_i}|$ i.e. all the absolute errors equal the same constant. It is straightforward to verify that CS is exact in this case.