A Differentiable Approximation to the Median

A. J. Allinger
Kruger Optical Corporation
15 September 2023

An approximation to the median of an array is developed. It is differentiable in the sense that if **y** is an array of functions of x, the derivative of the median of **y** may be written in terms of a sum of the partial derivatives $\frac{\partial y_i}{\partial x}$ weighted by a normal probability density.

Motivation. Many optimization problems aim to reduce the difference between estimated and target values. An analyst interested in optimizing the parameters of a machine learning system will typically aim to minimize a loss function such as the sum of the sum of the squares between the target values **y** and the estimates **ŷ** as in

$$f = \sum_i (\hat{y}_i - y_i)^2$$

It has often been argued that the sum of absolute differences is preferable, in part because it is more robust to outliers in the data. However, both the sum of squares and the sum of absolute values have a breakdown point of 0%, that is, a single spurious observation can cause an arbitrarily large change in the value of the loss f.

The median absolute deviation (MAD) is a highly robust estimate of dispersion. When attempting to cope with discrepant or dubious data, it is proposed to seek to minimize

$$f = \operatorname{med}_i |\hat{y}_i - y_i|$$

In order to apply the methods for optimization of continuous functions, it will be necessary to express the median of an array as a continuously differentiable function.

Condition on a quantile of a data set. Given an array **y** of length n, any quantile y_q must satisfy the counting function

$$\sum_i \begin{Bmatrix} -1 & y_i < y_q \\ 0 & y_i = y_q \\ +1 & y_i > y_q \end{Bmatrix} = (1-q)n - qn$$

For the case of the median, q is one half, and the right side of the equation reduces to zero. Rewriting the counting function with absolute values yields

$$\sum_i \frac{y_i - y_m}{|y_i - y_m|} = \sum_i \frac{|y_i - y_m|}{y_i - y_m} = 0$$

where y_m is the median.

An accurate approximation of the absolute value. A common approximation of the absolute value is

$$|u| \approx \sqrt{u^2 - \tau^2} - \tau$$

where τ is very small. A more accurate expression in terms of the error function has been proposed by Ravi. (2012)

$$|u| \approx u \operatorname{erf}\left(\frac{u}{\tau}\right)$$

Recently, this approximation has been proved superior to several alternatives. (Bagul and Chesneau, 2021)

The approximate median. Making use of this improved approximation gives for the median

$$\sum_i \operatorname{erf}\left(\frac{y_i - y_m}{\tau}\right) = 0$$

While an explicit analytical solution for y_m is not available, the equation is of a single unknown. Given a value of τ, it may be solved approximately by bisection over the interval y_{min} to y_{max} to any desired precision.

Derivative of the median. Suppose that y_i is an array of continuously differentiable functions of x, so that there exists an array of the partial derivatives $\frac{\partial y_i}{\partial x}$. Differentiating the equation for the median with respect to x yields

$$\frac{2}{\tau\sqrt{\pi}} \sum \exp\left(-\left(\frac{y_i - y_m}{\tau}\right)^2\right)\left(\frac{\partial y_i}{\partial x} - \frac{\partial y_m}{\partial x}\right) = 0$$

which may be solved explicitly

$$\frac{\partial y_m}{\partial x} = \frac{\sum \exp\left(-\left(\frac{y_i - y_m}{\tau}\right)^2\right)\left(\frac{\partial y_i}{\partial x}\right)}{\sum \exp\left(-\left(\frac{y_i - y_m}{\tau}\right)^2\right)}$$

The expression may be recognized as a linear combination of the partial derivatives, where the weights are the Normal density function for a suitable choice of scaling parameter τ.

Implementation. The computation of the approximate median has been implemented in `C` and FORTRAN languages and is freely available from the author's website. The usage is:

```
damed (double *y, double *dydx, int n, double tau, double *ymed, double *dmdx)
```

where `y` is the array, `dydx` is the array of partial derivatives, `n` is the length of the arrays, and `tau` is the scaling parameter. Results are `ymed`, the approximate median, and `dmdx`, the derivative of the approximate median.

Discussion. It has not been shown that this approximation of the median is optimal; it is possible that a better one exists. The scaling parameter τ must not be set too small. A value for τ of about 10% of the standard deviation is a reasonable first guess.

References.

"Smooth Approximation of absolute value inequalities", Ravi, 21 July 2012.
https://math.stackexchange.com/questions/172439/smooth-approximation-of-absolute-value-inequalities/173479#173479

"Sigmoid functions for the smooth approximation to the absolute value function," Yogesh J. Bagul and Christophe Chesneau, Moroccan Journal of Pure and Applied Analysis v.7 n.1 pp.12-19, 2021
https://intapi.sciendo.com/pdf/10.2478/mjpaa-2021-0002

Package DAMED
http://13olive.net/code/damed.zip