Online Weighted Mean

by Joshua Burkholder

Given the following set of inputs and their associated weights:



Let  be the number of inputs and their associated weights,  is the weighted sample mean for the first  inputs and their associated weights,  be the weighted sample mean of the first  inputs and their associated weights,  be the -th weight associated with input , and  be the -th input associated with . Then, the recurrence equation for the weighted sample mean (a.k.a. online weighted mean) is:



where . Preferably, all the weights are positive such that .

Proof:

The definition of the sample mean is:



The definition of the weighted sample mean is:



If we expand this definition, we have:



From algebra, we know that for arbitrary , , , and :



Hence, we have:



Since the weighted sample mean for the first  inputs and their associated weights is defined as  , we have:



Factoring out the , we have:



Combining the fractions and factoring out the , we have:



Therefore, the recurrence equation for the weighted sample mean (a.k.a. online weighted mean) is:



where .

Note: If all the weights are the same constant value  (i.e.  for ), the weighted sample mean would be:



For instance, if all the weights are , then the weighted sample mean is the sample mean:



Similarly, the online weighted mean with weights of the same constant value  would be:



Therefore, if all the weights are the same constant value , the online weighted mean is the same as the online mean.

Example of C++ code that computes the online weighted mean:

#include <iostream>

#include <iomanip>

int main () {

 double x;

 double weight;

 double sum\_of\_weights = 0;

 double weighted\_mean = 0;

 double prev\_weighted\_mean;

 if ( std::cin >> x && std::cin >> weight ) {

 sum\_of\_weights += weight;

 weighted\_mean = x;

 while ( std::cin >> x && std::cin >> weight ) {

 prev\_weighted\_mean = weighted\_mean;

 sum\_of\_weights += weight;

 weighted\_mean = (

 prev\_weighted\_mean - weight \* ( prev\_weighted\_mean - x ) / sum\_of\_weights

 );

 }

 }

 std::cout << "sum\_of\_weights: " << std::setprecision( 17 ) << sum\_of\_weights << '\n';

 std::cout << "weighted\_mean: " << std::setprecision( 17 ) << weighted\_mean << '\n';

}

Example of data.txt:

-19.313117172629575 2.718281828459045

-34.14656787734913 7.38905609893065

-14.117521595690334 20.085536923187668

. .

. .

. .

Command line:

g++ -o main.exe main.cpp -std=c++11 -march=native -O3 -Wall -Wextra -Werror -static

./main.exe < data.txt

Sample Output:

sum\_of\_weights: 34843.773845331321

weighted\_mean: -28.368899576339764