Pseudocode – k Nearest Neighbor

k-Nearest Neighbor

```
function knn(dataset, query, k)
    distances = []
    for example in dataset:
        add (distance(example, query), example) to distances
    end
    sort distances on first value in tuple.
    nearest = take k first values from distances
    return processing( nearest)
end
```

Comments

This is a very naive version of k-nearest neighbor. You may use it as a reference implementation or come up with something better in terms of computational cost. However, all things considered, a working program is better than an efficient one that doesn't work…as long as it finishes before the due date/time.

The processing() function could be used to implement something like weighted kNN or even to completely change the behavior of kNN to regression or classification by supplying a different processing function as a parameter. At that point, knn becomes a higher order function (HOF).

However, you only need to implement kNN regression for the programming assignment so if you don't understand all this…just ignore it. In that case, processing() is just returning the average value of y from the k nearest.

The distance() function could also be parameterized (more HOF!) but you need only use Euclidean Distance:

$$d(x, y) = (\sum_i (x_i - y_i)^2)^{1/2}$$