From youtube:

Very good explanation: <https://www.youtube.com/watch?v=H9H6s-x-0YE>

from wikipedia:

A **self-organizing map** (**SOM**) or **self-organizing feature map** (**SOFM**) is a type of [artificial neural network](https://en.wikipedia.org/wiki/Artificial_neural_network) (ANN) that is trained using [unsupervised learning](https://en.wikipedia.org/wiki/Unsupervised_learning) to produce a low-dimensional (typically two-dimensional) feature space. [wikipedia]

Self-organizing maps differ from other artificial neural networks as they apply [competitive learning](https://en.wikipedia.org/wiki/Competitive_learning) as opposed to error-correction learning (such as [backpropagation](https://en.wikipedia.org/wiki/Backpropagation) with [gradient descent](https://en.wikipedia.org/wiki/Gradient_descent)), and in the sense that they use a neighborhood function to learn and preserve the [topological](https://en.wikipedia.org/wiki/Topology) properties of the input space. [wikipedia]

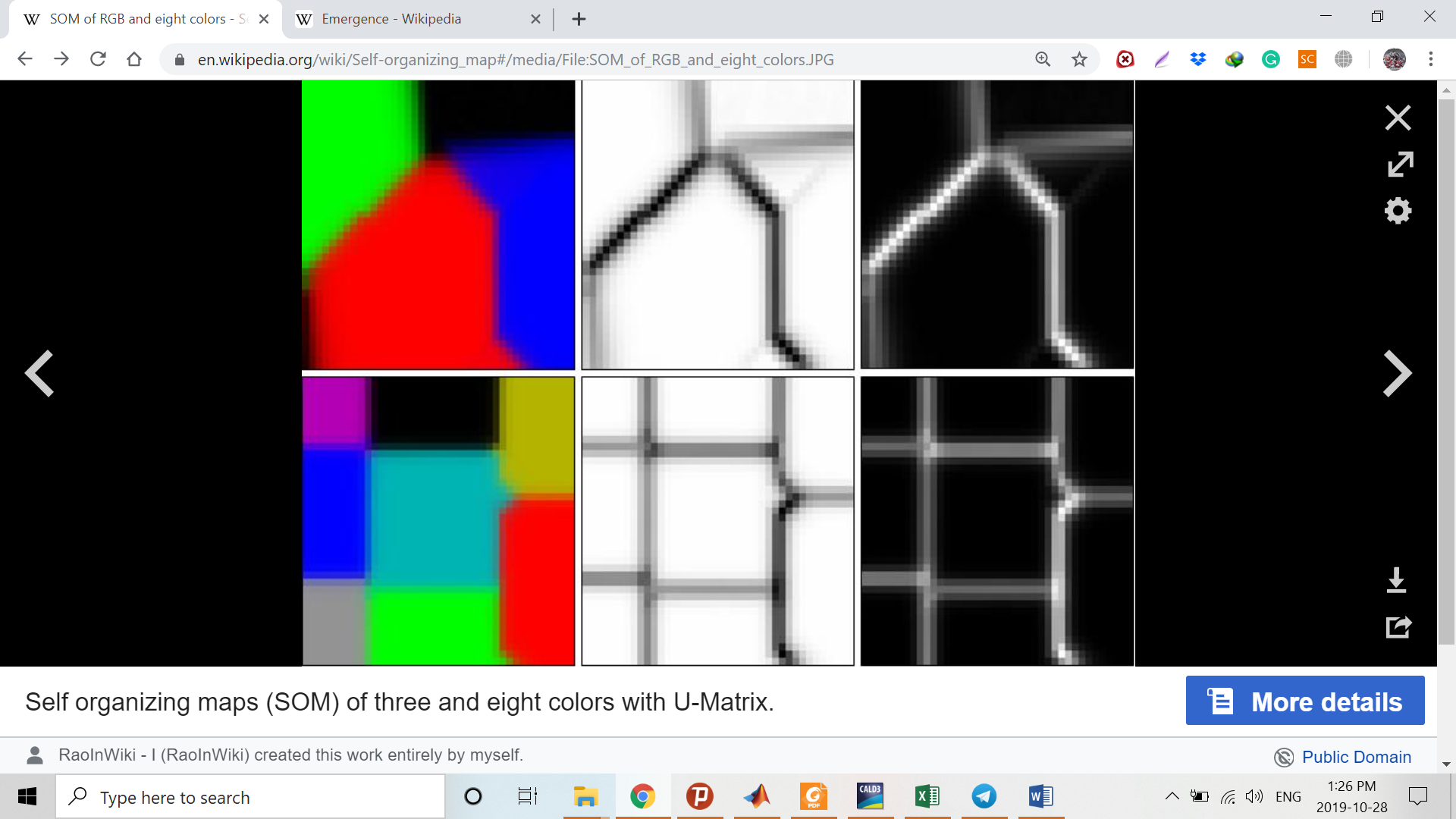
Useful extensions include using [toroidal](https://en.wikipedia.org/wiki/Torus) grids where opposite edges are connected and using large numbers of nodes. It has been shown that while self-organizing maps with a small number of nodes behave in a way that is similar to [K-means](https://en.wikipedia.org/wiki/K-means_algorithm), larger self-organizing maps rearrange data in a way that is fundamentally topological in character. [wikipedia]

It is also common to use the [U-Matrix](https://en.wikipedia.org/wiki/U-Matrix).[[5]](https://en.wikipedia.org/wiki/Self-organizing_map#cite_note-UltschSiemon1990-5) The U-Matrix value of a particular node is the average distance between the node's weight vector and that of its closest neighbors.[[6]](https://en.wikipedia.org/wiki/Self-organizing_map#cite_note-Ultsch2003-6) In a square grid, for instance, we might consider the closest 4 or 8 nodes (the [Von Neumann](https://en.wikipedia.org/wiki/Von_Neumann_neighborhood) and [Moore neighborhoods](https://en.wikipedia.org/wiki/Moore_neighborhood), respectively), or six nodes in a hexagonal grid. [wikipedia]

Like most artificial neural networks, SOMs operate in two modes: training and mapping. "Training" builds the map using input examples (a [competitive process](https://en.wikipedia.org/wiki/Competitive_learning), also called [vector quantization](https://en.wikipedia.org/wiki/Vector_quantization)), while "mapping" automatically classifies a new input vector. [wikipedia]

The training utilizes [competitive learning](https://en.wikipedia.org/wiki/Competitive_learning). When a training example is fed to the network, its [Euclidean distance](https://en.wikipedia.org/wiki/Euclidean_distance) to all weight vectors is computed. The neuron whose weight vector is most similar to the input is called the **best matching unit** (BMU). The weights of the BMU and neurons close to it in the SOM grid are adjusted towards the input vector. The magnitude of the change decreases with time (according to the learning rate parameter) and with the grid-distance from the BMU (according to the Euclidean distance). [wikipedia]

There are two ways to interpret a SOM. Because in the training phase weights of the whole neighborhood are moved in the same direction, similar items tend to excite adjacent neurons. Therefore, SOM forms a semantic map where similar samples are mapped close together and dissimilar ones apart. This may be visualized by a [U-Matrix](https://en.wikipedia.org/wiki/U-Matrix) (Euclidean distance between weight vectors of neighboring cells) of the SOM.[[5]](https://en.wikipedia.org/wiki/Self-organizing_map#cite_note-UltschSiemon1990-5)[[6]](https://en.wikipedia.org/wiki/Self-organizing_map#cite_note-Ultsch2003-6)[[16]](https://en.wikipedia.org/wiki/Self-organizing_map#cite_note-17) The other way is to think of neuronal weights as pointers to the input space. They form a discrete approximation of the distribution of training samples. More neurons point to regions with high training sample concentration and fewer where the samples are scarce. [wikipedia]



SOM may be considered a nonlinear generalization of [Principal components analysis](https://en.wikipedia.org/wiki/Principal_components_analysis) (PCA).[[17]](https://en.wikipedia.org/wiki/Self-organizing_map#cite_note-18) It has been shown, using both artificial and real geophysical data, that SOM has many advantages[[18]](https://en.wikipedia.org/wiki/Self-organizing_map" \l "cite_note-19)[[19]](https://en.wikipedia.org/wiki/Self-organizing_map#cite_note-20) over the conventional [feature extraction](https://en.wikipedia.org/wiki/Feature_extraction) methods such as Empirical Orthogonal Functions (EOF) or PCA. [wikipedia]

