Chapter 13 : Part 1

**Deep Learning**

**SOM: Self-Organizing Maps**

Introduction

**13.1.1 What we will learn in this Chapter**

1. How do Self-Organizing Maps work?: First of all we will talk about how *self-organizing maps (SOMs)* work. It will help us understand what to expect, what we're aiming for. We'll know the end goal that we're working towards.
2. K-Means Clustering: Then we'll talk about *K-Means Clustering*, it will be a review for us what we've done in ML in Chapter 4: Clustering.
3. How do Self-Organizing Maps Learn? (Part 1 & Part 2): We'll talk about how do *self-organizing maps* learn, in these 2-part. We do this in two parts because we'll dive deep into the topic here.

* We'll walk through them step by step example for better understanding.

1. Live SOM example: Here we'll have a live SOM example, a very simple one. Here you will see how a SOM *structures itself* and preserves *similarities* & *correlations* in your data set and portrays them in a *lower dimensionality* representation (*2-D map*).
2. Reading an Advanced SOM: Finally we will talk about reading an *advanced SOM*. This shows you how to read those *SOMs*.

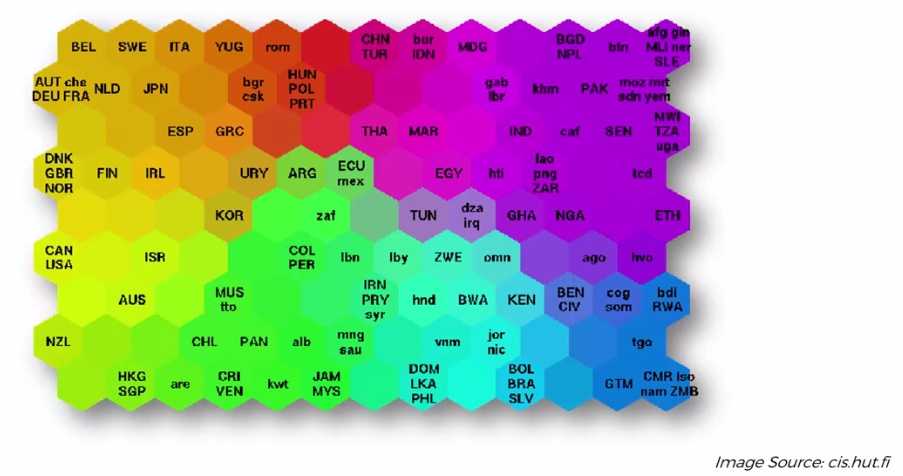
* We'll have different maps on one screen and by looking at them you can *read/understand* them.
* We'll discuss some examples of *map* *implementations* to guide you in the direction of where you can do *further study* in the space of *SOMs*.

**13.1.2 SOM (Self-Organizing Map)**

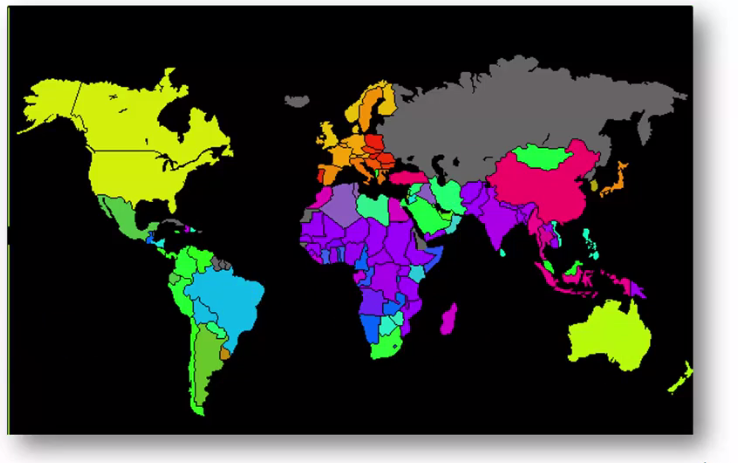
|  |  |  |
| --- | --- | --- |
| We already talked about ANN, CNN and RNN those are collectively called supervised deep learning.   * However, SOM is a unsupervised deep learning method. *Self-organizing maps (SOM)* were invented in the 1980s by, *Teuvo Kohone*. * Sometimes *SOMs* even called the *Kohonen* *maps*. | |  |
| * Usage of SOMs: * SOMs are used for *reducing* *dimensionality*. * SOMs can be used in *astronomy*. Here's a great example from the paper. |  | |

Here we got a beautiful visualization of how *self-organizing maps* actually work.

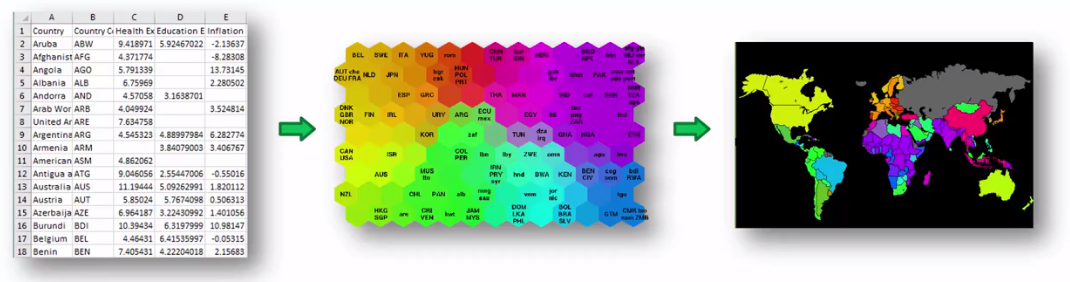
* SOMs take a *multi-dimensional data set* (with *lots* of *columns* which are the *dimensions* of the data set, and *lots* of *rows*) and they *reduce* the *dimensionality* of these *data* *sets*.
* So basically, instead of having ***20***, ***30*** or a ***hundred*** or even more ***columns*** (20, 30 or 100 dimension), you end up with a ***map***.
* That's why they called Self-Organizing Maps (SOMs) because we end up with a *two-dimensional representation* of your *data* *set*.
* The purpose of the SOM is to *reduce* the amount of *columns*. And represent the data into 2D-map.
* Example: Here is an actual SOM. which was produced, from the *data* of the different *states* of *prosperity* and *poverty* in different *countries*.



* Those names actually represents *countries* of the world and this *SOM* has put them into *clusters* based on lots of *different* *indicators*.
* In this specific example, 39 different indicators were used. And indicators are parameters describing things such as *quality of life*, *factors*, the *state* *of* *health* in a country, *nutrition*, *educational* *services*, and so on.
* We can see that in the top left corner, we have *countries* with the *best* or the *least alarming* state of *poverty*. Those countries are ***Belgium***, ***Sweden***, ***Japan***, ***Spain***.
* We also notice that it's slowly going towards the other end of spectrum where you have countries with the most *alarming* state of *poverty*, like *Ethiopia* or *Zimbabwe*.
* So if we have a huge data set containing 200 plus *countries* as *rows*, and 39 different columns (dimensions. Indicators in SOM). So it's impossible to visualize.
* But using a SOM, we can *reduce* the *dimensionality* and present it as a *map* like above.
* Remember that, SOMs are unsupervised techniques: It has *training* *data* but it doesn't have any *labels* in the *training* *data*. So its learning on its own.
* Basically just given data, and then it learns to group these data (countries). It is much more like clustering in ML.
* What else can we do with this map: We take the *color-codes* for the different *countries* and *color* them in the *world* *map*. In this world map we can determine first world countries and third world countries, and where countries are developed, where countries are still developing.



* To summarize, we had the data, which you can get from the *World* *Bank* (you can just download data sets from there),
* We *reduce* the *columns* using *SOM* and get a *2D-map* with different *countries* *grouped* *together* according their *color* *code*. It would help you group your data set. SOM map is still a good representation for the data but we can use the colors in the world map.
* Next, we can color the countries in world map using SOM map.



* SOM can be applied to visualize different kind of data: It groups the data, so that you would understand different similarities based on all of your data. You wouldn't have to go through hundreds and hundreds of columns.
* Grouping different types of *equipment* that you might be considering or you might be *selling* through your organization.
* Grouping different types of *stock* and *inventory*.

You would be able to just look at this map and quickly understand all of the similarities.

|  |  |
| --- | --- |
| * Additional reading: If you'd like to get some additional reading a good paper to check out by Teuvo Kohonen from 1990, it's called *"The Self-Organizing Map"*. |  |