

# Picking a Dataset :

200 total songs, evenly distributed across rap, reggae, classical, and country genres

The musical style were chosen because they were supposed to be easy to tell apart. It might be a necessary compromise to find success in this exercise, but we can wonder if it doesn't undermine the efficiency in practical use : people rarely listen to only 5 music genre which are all very different, and a database in general has no reason to contain only a few genres which are usually not listened to by the same public.

An online radio LastFM provided its users specific playlist based on the user's favorite genres. They made

## PREPROCESSING :

Features : Eight summary features about each song

Time Signature, Energy ,Tempo , Mode, Key, Duration, Loudness, Danceability

The features are calculated with the help of Echonest, an online music library. They did not extract their own features, which is probably one of the hardest part of the job.

### First PROCESSING

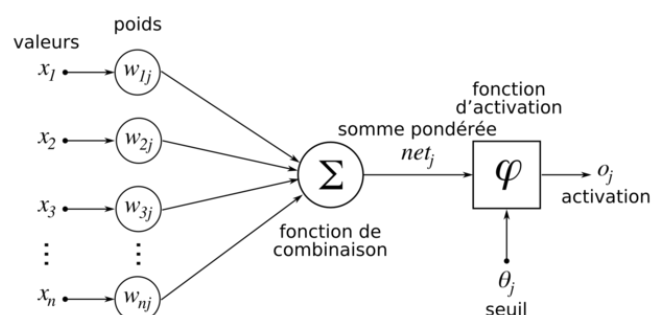
They used a combination of a growing neural gas and a neural network, as these methods are supposed to be complementary.

### Growing neural gas :

Each vector is added to a « map », and attributed a « node ». A node regroups vector which are close enough, if a new vector is too far from all existing node, it creates a new node. This process reduces noise and the next step classifies the nodes (which are the average of the vectors of each nodes), instead of all the initial vectors, which is faster.

### Neural network

Each row of neurone progressively sets its weight to get the best results on the training set. The training phase is very costly in terms of tim, and the results greatly depends on the



activation function (which can be any function that takes a real number and gives a real between 0 and 1)

They used these first steps to engineer new data for each song, they picked the two closest node of GVG, and the confidence values for those vectors (which depends on the distance between said nodes and the vector representing the song).

With this information they decided to train the neural networks on seven different sets. 1. song attributes (8 inputs) 2. model vector (8 inputs) 3. model vector and song (16 inputs) 4. model vector and song and confidence value (17 inputs) 5. 1st and 2nd model vectors (16 inputs) 6. 1st and 2nd model vectors and song (24 inputs) 7. 1st and 2nd model vectors and song and confidence value (25 inputs).

### **Results**

With their setting for GNG, they got 12 nodes on average (the nodes creation depends on the order of the songs).

Schematically, they got on average 3 nodes reggae, 2 rap, 3 country, and 4 classical.

The best results were obtained using the song vector and its two closest nodes. Adding the confidence values reduced accuracy despite all their tries.

Final accuracy : 88% on training set, 68% on testing set

Advantages of the method :

It is possible to check if the GNG works as expected, by looking at the nodes populations and seeing if similar genres are closer, and if vectors of a same node tend to be representing the same genre.

It's not a « boîte noire », which helps seeing which feature is relevant without having to train the full program.

Few features (7 if we remove duration) = reasonable complexity

Possible ameliorations :

Neuro Evolution of Augmenting Topologies (NEAT), a modification of a neural net that has a dynamic hidden layer. More specifically the system only introduces nodes to the hidden layer when they are needed. It would help make the system much easier to adapt to new sizes of vector, or different numbers of musical genre

Comments :

Their Data set seems a bit small, and the extraction of vector is rather lazy (only 8 attributes, we don't know how they are calculated, they did not use portions of the song, which seems to work well)