Project Proposal

Amlan Kar, Chaitanya Ahuja

2015-03-16 Mon

Abstract

Music Classification is a harder task as compared to speaker classification, due to presence of polyphonic sounds in the in signal. Hence signal-processing techniques fail to produce usable competent results. We attempt to deploy a deep neural network for extracting features [1] from different genres of music. This system would be then used for classyfying unknown music signals. As opposed to a traditional neural net, we would use Rectified Linear Units (ReLUs) as the activation function and a random forest classifier for classification.

Flowchart

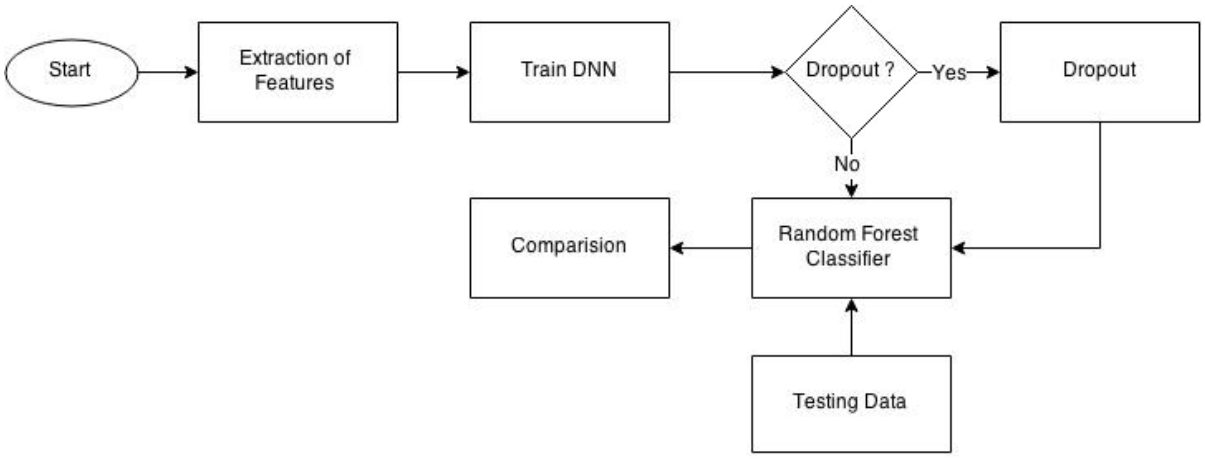

Figure 1: **Flow of algorithm**

Aim

The broad aim of this project is a better extraction of features of a music signal using Deep Neural Networks and scale this system up for large training sets. Along with that we would use this trained network as a classifier to classify music signals into different genres such as rock, jazz, classical etc. As a starter, we would broadly go in line with the following steps:

- Creating a training and testing database randomly on a labeled set.

[1] Sigtia, Siddharth, and Simon Dixon. "Improved music feature learning with deep neural networks." Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on. IEEE, 2014.

- Train the neural using the ReULs (Rectified Linear Units), instead of the traditional Sigmoid function, by running the SGD (Stochastic Gradient Descent) on the complete network.
- Find the optimal learning rate and dropout rate using a grid-search and hence train the neural net with the optimal findings.
- For classification, use a random forest classifier on the activations of the hidden layer as features and hence create a model for classifying the genres of the test music sample.

Introduction to the Key concepts

- **DNN:** Deep learning is an attempt to model abstractions in data using architectures composed of multiple non-linear transformations. A deep neural network is an artificial neural network with multiple hidden layers of units between the input and output layers. DNN's are used to model complex non-linearities in data. Deep architectures generate models where an object can be expressed as a layered composition of primitives, as can be seen in the image below.

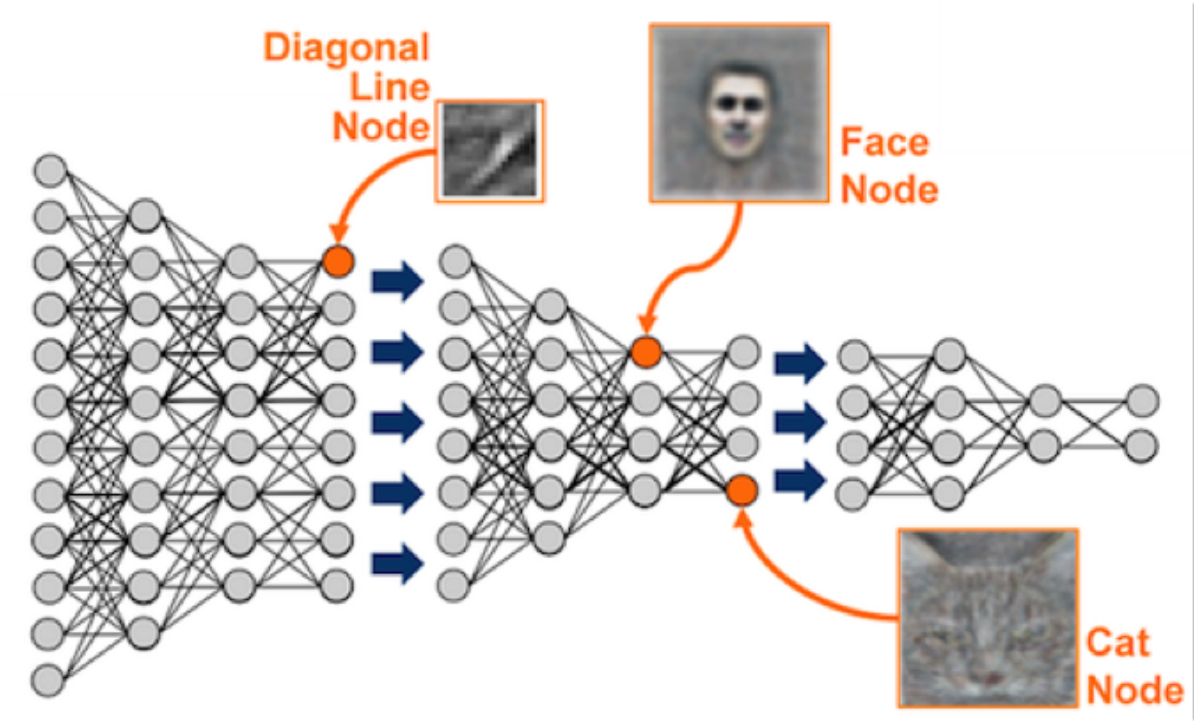

Figure 2: **A simple DNN**

- **RNN:** A recurrent neural network is another type of an artificial neural network, where each unit contains a directed edge to every other unit, thus creating directed unit cycles. When operated in discrete time settings (time-varied real valued activation function), these systems can capture complex long-term dependencies between its inputs.

- **Dropout:** Deep Neural Networks, by virtue of their multiple non-linear hidden layers can learn various complicated relationships from data. But most of the time, when the training set is small, these complications arise from the sampling noise, thus resulting in overfitting. Dropout is a powerful

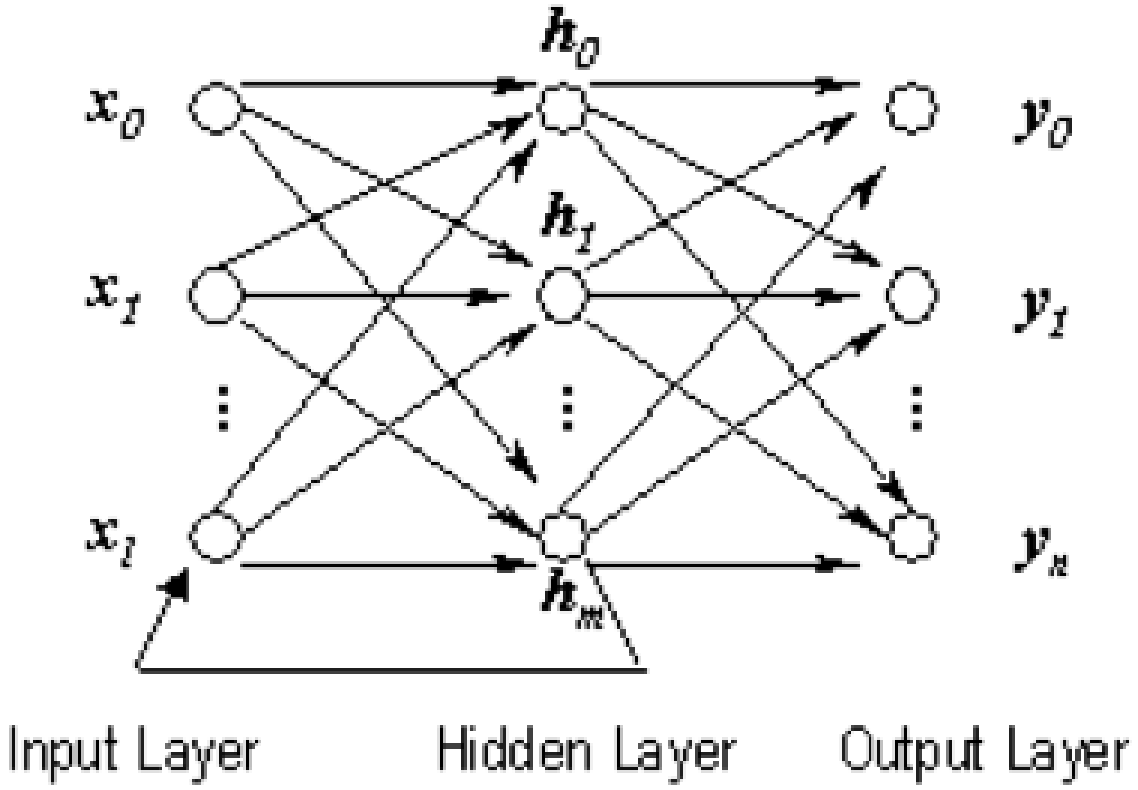

Figure 3: **A simple RNN**

regularization method [2], used to prevent DNN's from overfitting. The term "dropout" refers to dropping out units (hidden and visible) in a neural network. Dropping an unit out is simply removing it from the network, along with all its incoming and outgoing connections. The choice of which units to drop is random. In the simplest case, each unit is retained with a fixed probability p independent of other units, where p can be chosen using a validation set or can simply be set to a certain value.

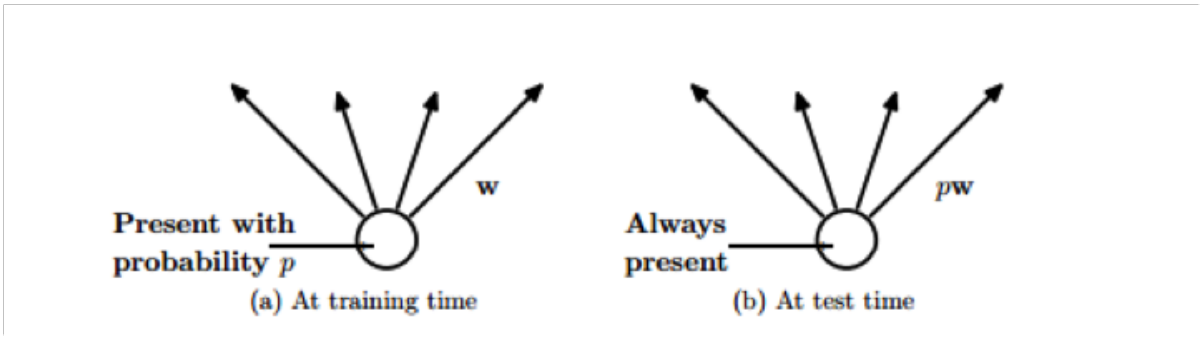

Figure 4: **Dropout Example**

- **Rectified linear units:** Rectified Linear Unit (ReLU) function is an activation function defined as max(0,x), where x is the input to a neuron.

[2]Srivastava, Nitish, et al. "Dropout: A simple way to prevent neural networks from overfitting." The Journal of Machine Learning Research 15.1 (2014): 1929-1958.