MUSIC CLASSIFICATION USING DNNS

Course Project for CS365

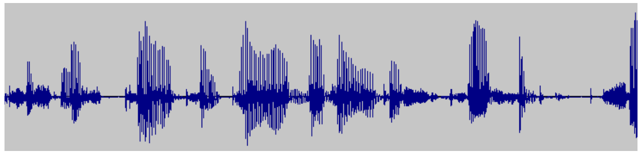
Chaitanya Ahuja

Amlan Kar

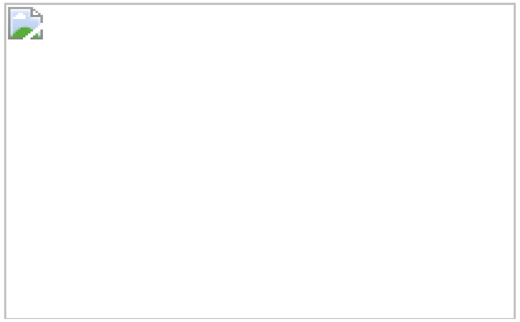
Mentored by Prof. Amitabh Mukherjee

PROBLEM STATEMENT

Music ⇨ Model ⇨ Artists/Genre



http://www.wirelesscommunication.nl/reference/images/voicesig.gif



http://img0.gtsstatic.com/wallpapers/a465cc841c36511acc5a3a3655795d40_large.jpeg

MODEL

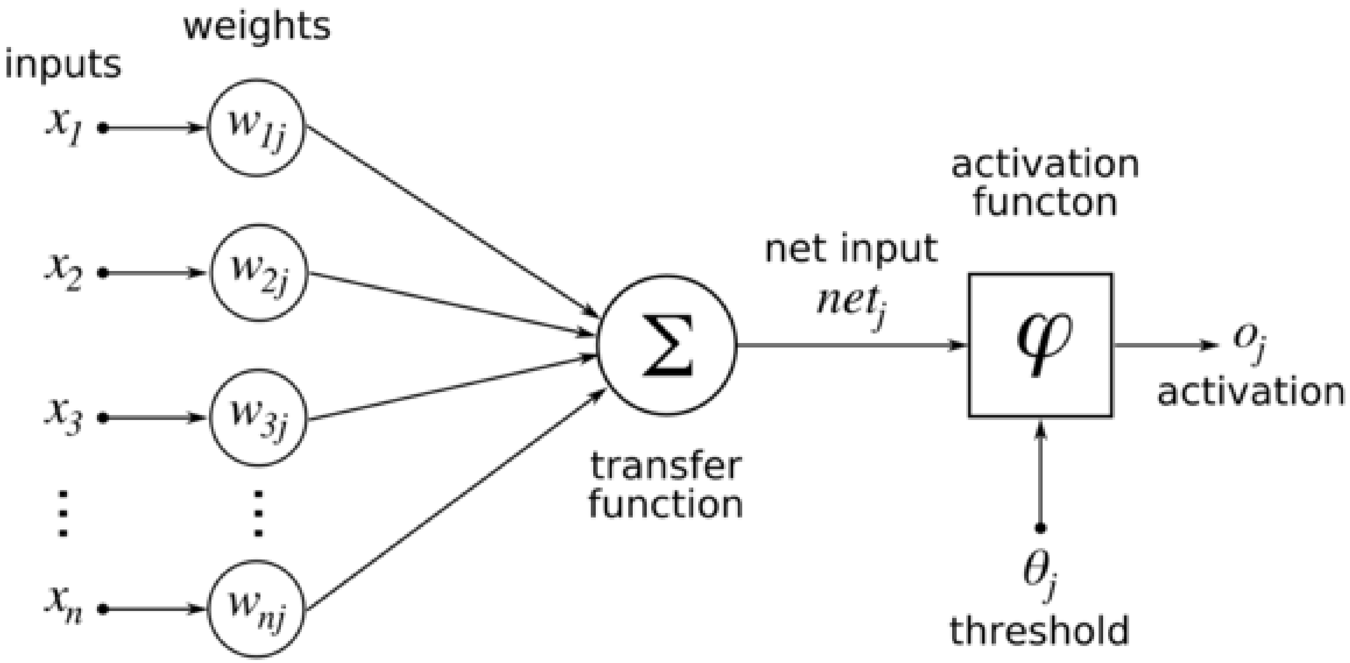
FEATURES

- Handcrafted
  - FFT
  - Cepstrum
  - MFCC
- HMM
- Neural Nets

CLASSIFIER

- Random Forests
- Neural Nets

NEURAL NETS

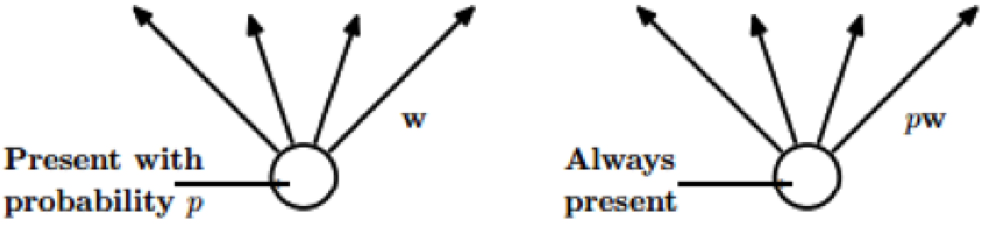

weights
inputs
x_1
w_{1j}
x_2
w_{2j}
x_3
w_{3j}
x_n
w_{nj}
activation
functon
net input
net_j
Σ
φ
o_j
activation
transfer
function
θ_j
threshold

WHY NEURAL NET FEATURES ?

- Have shown to work well for random weights in the DNN structure.
- Any set of features can be well learnt in a DNN setting
- DBNN features give advantage over hand-crafted features

DROPOUT

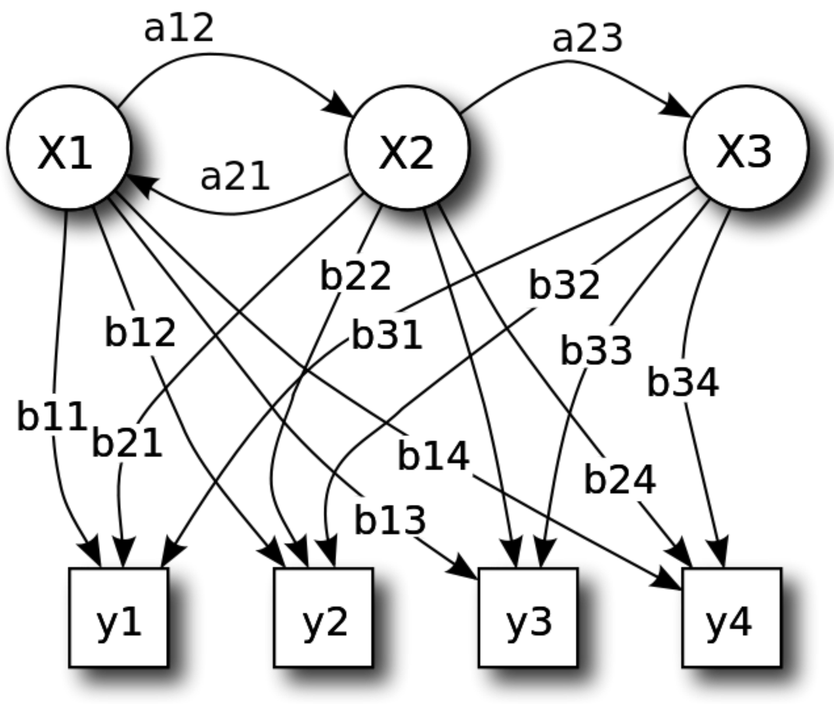
The term “dropout” refers to dropping out units (hidden and visible) in a neural network.



(a) At training time (b) At test time

HIDDEN MARKOV MODELS

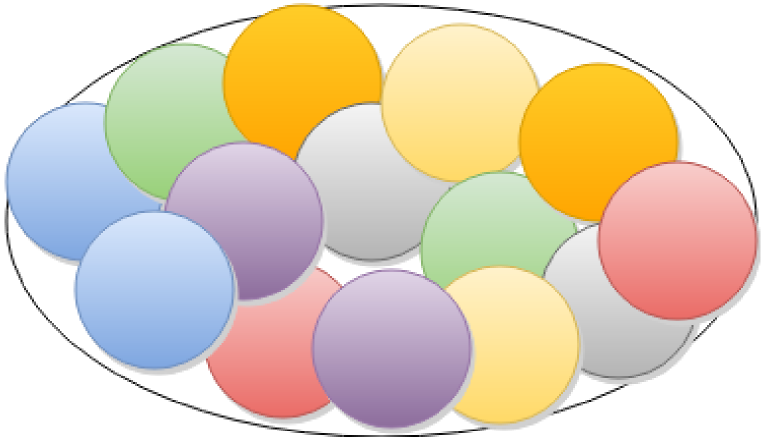
- A state-space model of the given form
- Takes data points sequentially as states and trains the weights accordingly
- Each state generates a probability distribution over the outputs
- Incorporates temporal information and hence works great with speech and music



*picture taken from wikipedia.org

CLASSIFICATION

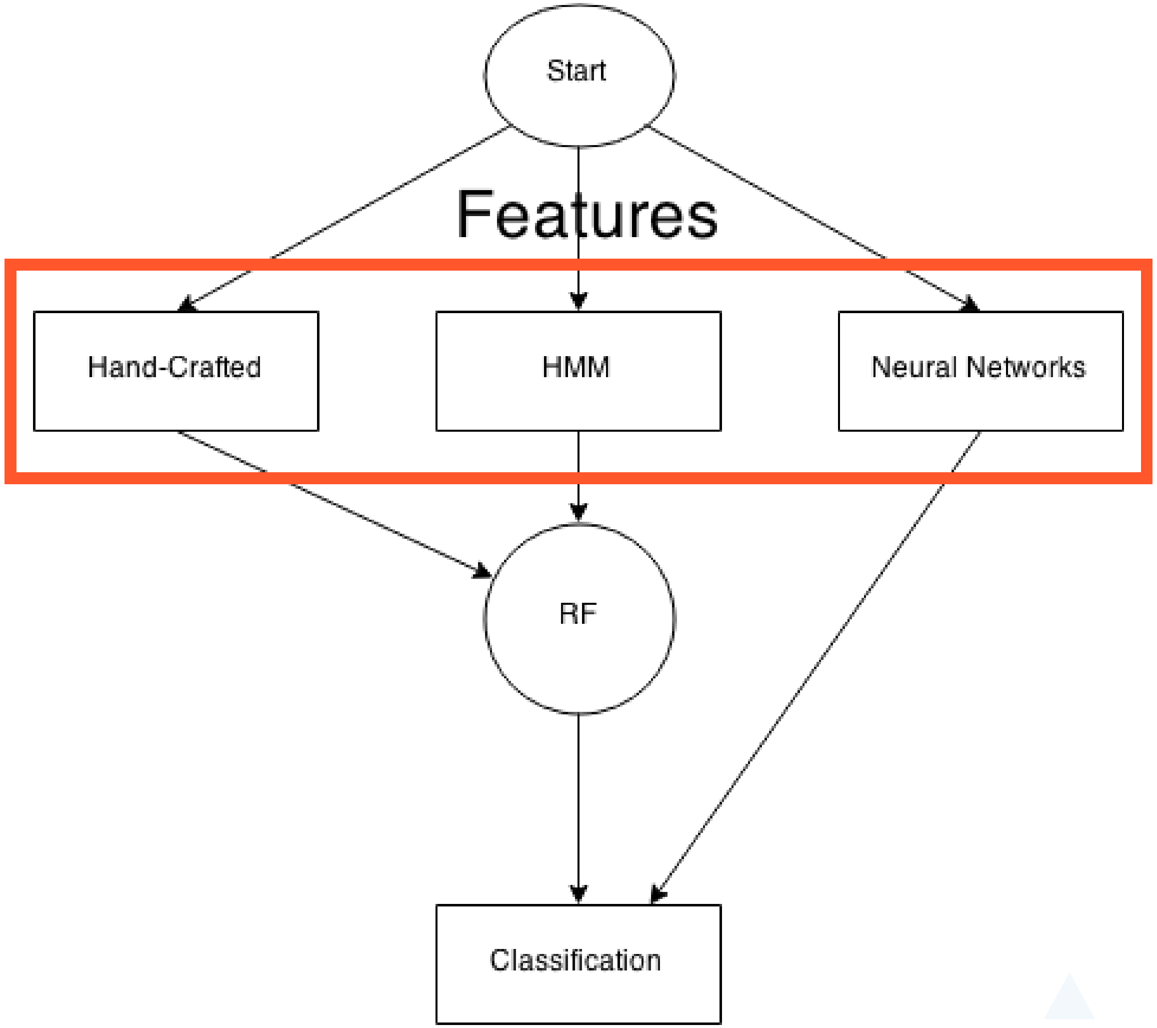
Random Forest (RF) classifier



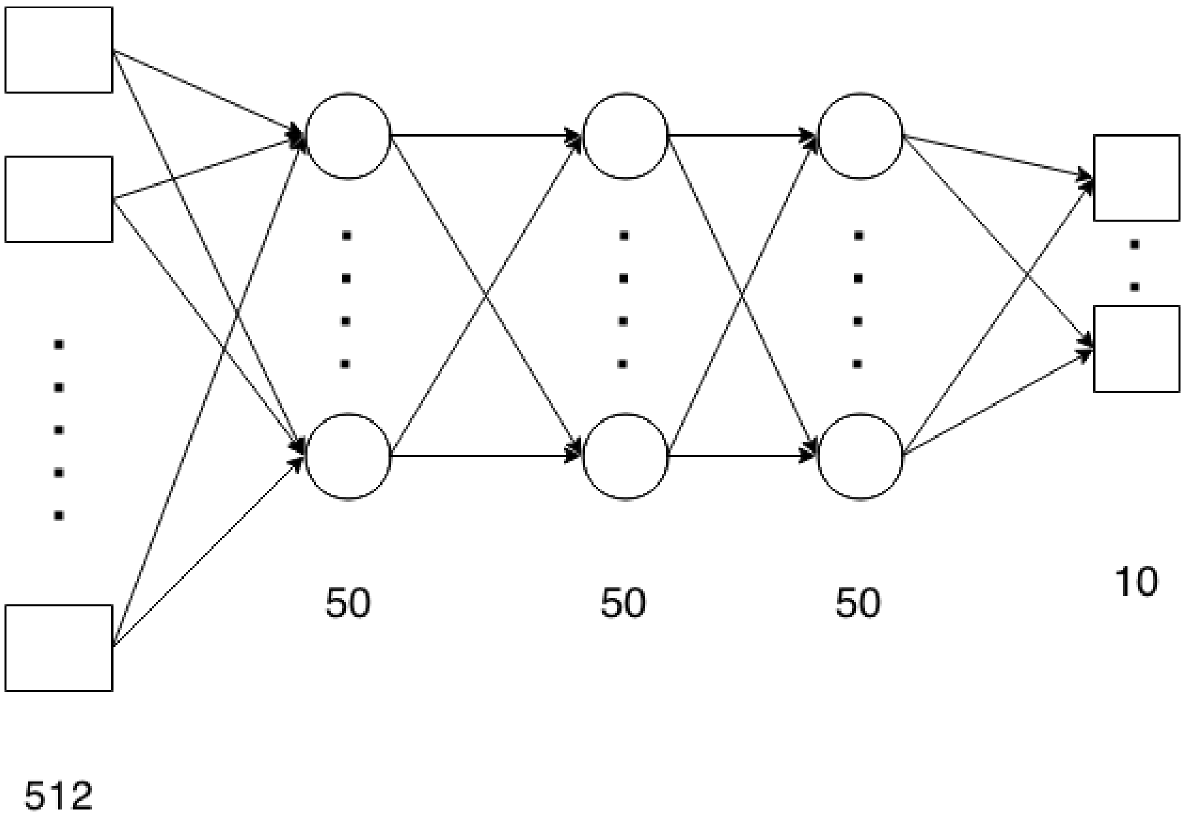
Why RF classifier over NN classification ?

- RFs do not overfit as compared to a typical DNN
- RFs can classify non-metric spaces

FLOWCHART


Start
Features
Hand-Crafted
HMM
Neural Networks
RF
Classification

NEURAL NETWORK STRUCTURE



RESULTS

- Training completed for genre classification (weights and activation values obtained)
- Need to test on test data to check results
- Here cost 0 is the loss function value at the input, cost 1 is the accuracy on the validation set. The maximum validation accuracy achieved in 50 epochs was 0.62
- Training with more epochs (the paper used 500) should give much better results
- Sigmoid function has been used as the output mask for each node

```
  Epoch 44
***Train Results***
Cost 0: 2.537269
Cost 1: 0.557686
***Validation Results**
Cost 0: 2.582240
Cost 1: 0.575616
  Epoch 45
***Train Results***
Cost 0: 2.535966
Cost 1: 0.556506
***Validation Results**
Cost 0: 2.604529
Cost 1: 0.592494
  Epoch 46
***Train Results***
Cost 0: 2.533423
Cost 1: 0.555090
***Validation Results**
Cost 0: 2.559155
Cost 1: 0.568234
Best Params!
```

What Next?

- Perform unsupervised learning to Deep Belief networks to get a better feature set
- Compare results obtained from features of DNN, DBN and HMM

References

- Saxe, Andrew, et al. "On random weights and unsupervised feature learning."*Proceedings of the 28th International Conference on Machine Learning (ICML-11)*. 2011.
- 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.
- Gales, Mark, and Steve Young. "The application of hidden Markov models in speech recognition." *Foundations and Trends in Signal Processing* 1.3 (2008): 195-304.
- Hamel, Philippe, and Douglas Eck. "Learning Features from Music Audio with Deep Belief Networks." *ISMIR*. 2010.
- 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.