Comparative Analysis of Classifiers on Audio, Text, and Image Datasets

Akshay Rana, Harmanpreet Singh, Himanshu Arora

Université de Montréal / Mila

OBJECTIVE

To explore classification performance of popular machine learning algorithms and deep learning algorithms on Audio, Image, and Text datasets.

Investigate the versatility of Naive Bayes, Support Vector Machine (SVM), Logistic Regression, MLP, Convolutional Neural Network and LSTM Networks.

ALGORITHMS

Naive Bayes classifier assumes all the features to be conditionally independent and hence can be extremely fast even on a high-dimensional distribution. In spite of overly simplified assumptions, it seems to perform quite well on real-world situations like text classification problems.

Logistic Regression is a linear classification method that learns the probability of a sample belonging to a certain class. We also plan to examine regularization techniques to avoid overfitting, resulting in a more generalized model.

Support Vector Machine is one of the most powerful classifiers that we will use in all our datasets. It is very effective in high-dimensional space and it can create both linear and non-linear decision boundaries.

Multilayer Perceptron can learn a complex non-linear function approximator using multiple hidden layers and is a great fit for our classification problems. We will do grid-search to explore the optimal hyper parameters like number of neurons, hidden layers, activation functions, etc.

Conv Nets & LSTMs are deep neural network architectures that have achieved state-of-the-art results on image and text classification problems. These proved to be useful in our use case as we have 2D images and sequential audio data.

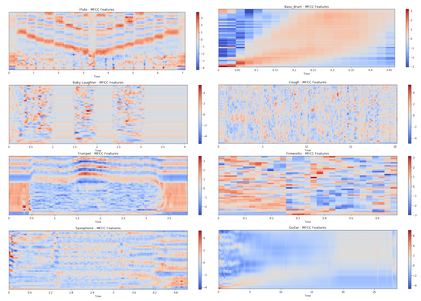
DATASET

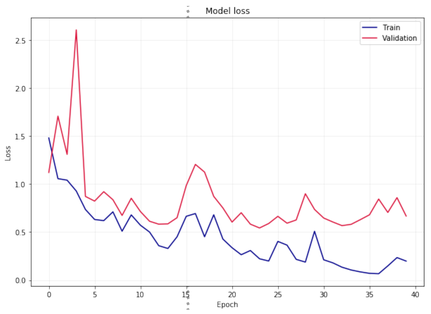
Audio: We conducted experiments on FreeSound dataset 2018 to automatically recognize sounds from a wide range of real-world environments. The task setup is a multi-class classification problem. We used spectral frequency-based methods for feature extraction.

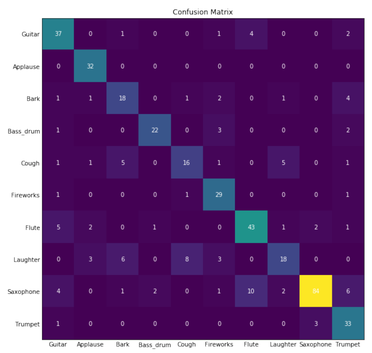
Text: We explore Sentiment140 dataset for sentiment classification of tweets. We explored machine learning as well as deep sequence-learning based classification methods along with word embeddings for feature extraction.

Image: For image classification, we investigated the CIFAR-10 dataset. This dataset consists of 60,000 color images of size 32x32 evenly distributed in 10 classes. We present our results on machine learning as well as CNN based models.

AUDIO: FREESOUND

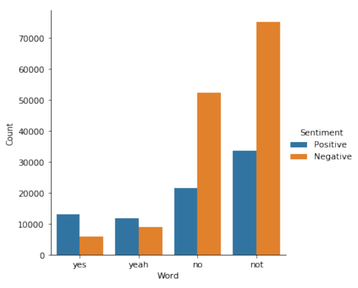


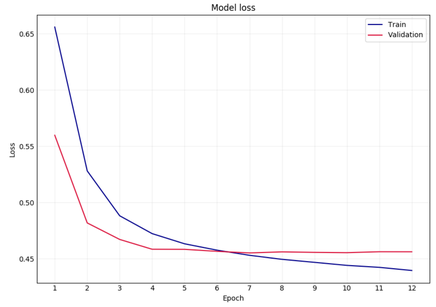


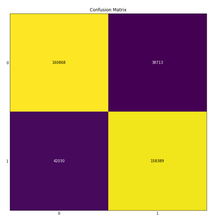


FEATURE DIM: 40 X 173 **TRAIN SIZE: 2882**
CLASSES: 10 **TEST SIZE: 435**

TEXT: SENTIMENT 140

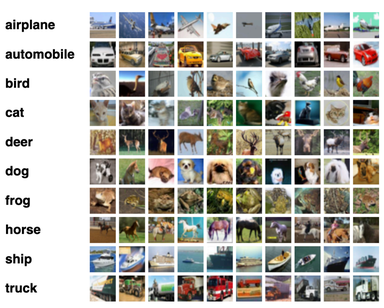


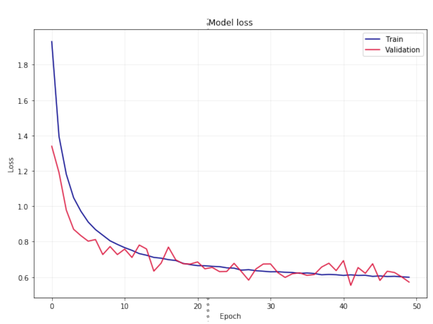


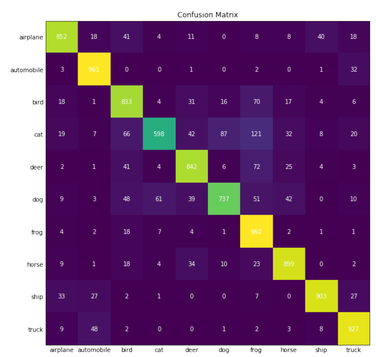


FEATURE DIM: 53 X V **TRAIN SIZE: 900k**
CLASSES: 2 **TEST SIZE: 400k**

IMAGE: CIFAR 10







FEATURE DIM: 32 X 32 X 3 **TRAIN SIZE: 50k**
CLASSES: 10 **TEST SIZE: 10k**

RESULTS

	Audio FREESOUND 2018		Text SENTIMENT140		Image CIFAR - 10	
Algorithms	Train	Test	Train	Test	Train	Test
Naive Bayes	.56	.55	.85	.80	.29	.28
Logistic Regression	.46	.40	.86	.81	.30	.26
SVM	.90	.54	**.91**	**.81**	.36	.29
Multi Layer Perceptron	.31	.32	.81	.80	.48	.50
CNN	.96	**.81**	-	-	.87	**.86**
LSTM	-	-	.80	.78	-	-

CONCLUSIONS

- Different algorithms work well for specific tasks and specific data modalities
- Convolutional Neural Networks work better for spatial datasets such as images and audios
- Textual data is generally best modeled by the Naïve Bayes classifier and recurrent neural networks
- Naïve Bayes classifier handles curse of dimensionality well.
- Task-dependent pre-processing of the data is important to get high accuracy.
- Careful review of algorithms prior assumptions is required when choosing which one must be used.
- There is no single algorithm that can be used in every situation with good results.
- Dense feature representation is crucial for text and audio classification.
- Transfer learning can be used for all three modalities to improve scores.
- LSTM based models can be explored for audio classification considering it as sequence data.

REFERENCES

- Let's keep it simple, Using simple architectures to outperform deeper and more complex architectures. *Seyyed Hossein Hasanpour, Mohammad Rouhani, Mohsen Fayyaz, Mohammad Sabokrou*
- On stopwords, filtering and data sparsity for sentiment analysis of Twitter. *Saif, Hassan; Fernández, Miriam; He, Yulan and Alani, Harith (2014)*
- CNN Architectures for Large-Scale Audio Classification. *Shawn Hershey et. al*