



Awake Networks Challenge

Ramkishore Swaminathan



Overview

- Problem Statement
- Data Preprocessing
- Intuition
- Multi Layer Perceptron
- LSTM

Problem Statement

- FizzBuzz
 - Divisible by 3 - Fizz
 - Divisible by 5 - Buzz
 - Divisible by 15 - FizzBuzz
- Build a Machine Learning Model to generate FizzBuzz for numbers from 1-1000
 - Features of ML model cannot test divisibility by 3 or 5
 - Training Data
 - 800 random numbers between 1 and 1000
 - Testing Data
 - Remaining 200 numbers

Data Preprocessing

- Input Representation
 - Fix number of digits based on maximum number - (1000)
 - Convert numbers to binary representation
 - Prepend with 0s if less than 10 digits
 - Eg: Representation of 3 will be 0000000011
- Output Class Representation
 - One Hot Encoding
 - FizzBuzz (0, 0, 0, 1)
 - Buzz (0, 0, 1, 0)
 - Fizz (0, 1, 0, 0)
 - Others (1, 0, 0, 0)

Intuition

- Divisibility cannot be used as a feature
 - Most obvious feature
- Not possible to hand-pick features for numbers
- Utilize Supervised Feature Learning
 - Relationship between input and output learnt by the model on its own
- Multilayer Perceptron (Feed Forward Neural Network)
- LSTM (Recurrent Neural Network)

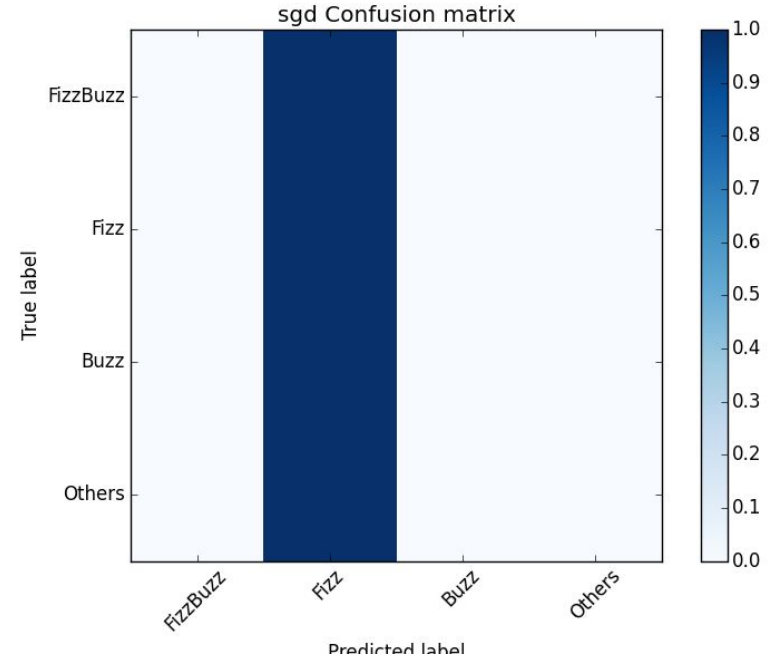
Multi Layer Perceptron

- Combination of Hyper Parameters experimented
 - # of Hidden Layers 1, 2, 3, 4, 5
 - # of Hidden Nodes 100, 200
 - # of Epochs 10000
 - Batch Size 100
 - Dropout 0, 0.5
 - Optimizer RMSProp, AdaGrad, SGD
 - Activation Function tanh, ReLU
 - Cost Function Softmax Cross Entropy
- None of the combinations were able to make out the differences between the different classes
 - Every model predicted one class for all the test instances as shown in the confusion matrix - (next slide)

Multi Layer Perceptron

Model Configuration

- 5 Layers
- 100 Hidden Nodes
- 10000 Epochs
- 100 Batch Size
- No Dropout
- SGD Optimizer
- tanh Activation
- Softmax Cross Entropy Cost Function



LSTM

- Long Short Term Memory - Recurrent Neural Networks
 - Learn long term dependencies
 - Remove or add information to the Cell state based on gates
- Combination of Hyper Parameters experimented
 - # of Hidden Layers 1, 2, 3, 4, 5, 6
 - # of Hidden Nodes 100, 200, 300
 - # of Epochs 10000
 - Batch Size 100
 - Dropout 0, 0.25, 0.5
 - Optimizer RMSProp, AdaGrad, SGD
 - Activation Function tanh, ReLU
- LSTM outperformed Multi Layer Perceptron by a huge margin

LSTM

Model Configuration

- 5 Layers
- 300 Hidden Nodes
- 10000 Epochs
- 100 Batch Size
- 0.25 Dropout
- RMSProp Optimizer
- tanh Activation
- Softmax Cross Entropy Cost Function

Accuracy - 100%

