



# Deep Learning Final Project LSTM Analysis

Alptekin Orbay 2017700090



# Motivation

- Recent studies proposing simple models
- Understanding gates functionality
- Problem specific LSTM types
- More stable and stronger hidden states

# Proposed LSTM TYPES

- Basic LSTM
- Only Forget
- Forget-Output
- My Forget-Output
- Input-Output
- Output

## Only Forget

$$f = \sigma(W_f * x_t + R_f * h_{t-1} + b_f)$$

$$z = \tanh(W_z * x_t + R_z * h_{t-1} + b_z)$$

$$c_t = c_{t-1} * f + z * (1 - f)$$

$$h_t = c_t$$

# Forget-Output

$$f = \sigma(W_f * x_t + R_f * h_{t-1} + b_f)$$

$$z = \tanh(W_z * x_t + R_z * h_{t-1} + b_z)$$

$$o = \sigma(W_o * x_t + R_o * h_{t-1} + b_o)$$

$$c_t = c_{t-1} * f + z * (1 - f)$$

$$h_t = \tanh(c_t) * o$$

# My Forget-Output

$$f = \sigma(W_f * x_t + R_f * h_{t-1} + b_f)$$

$$z = \tanh(W_z * x_t + R_z * h_{t-1} + b_z)$$

$$o = \sigma(W_o * x_t + R_o * h_{t-1} + b_o)$$

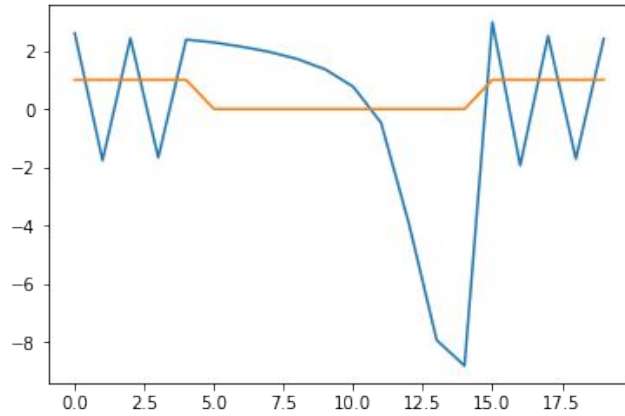
$$c_t = f * (c_{t-1} + z)$$

$$h_t = \tanh(c_t) * o$$

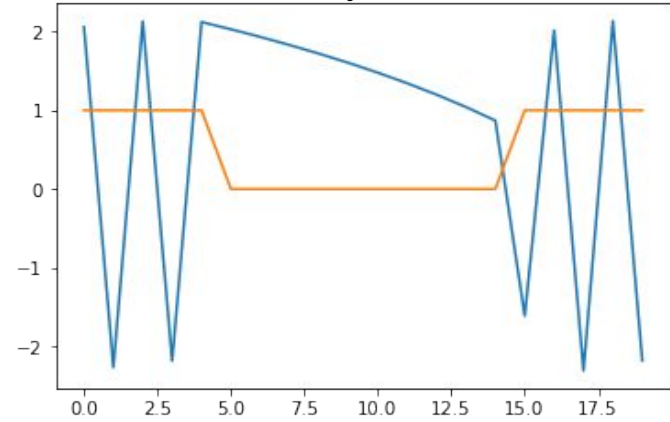
# Parity Check

- Artificial Dataset
- Hard to learn
- 1 Hidden Unit for Visualization
- Hard decisions(Saturation)

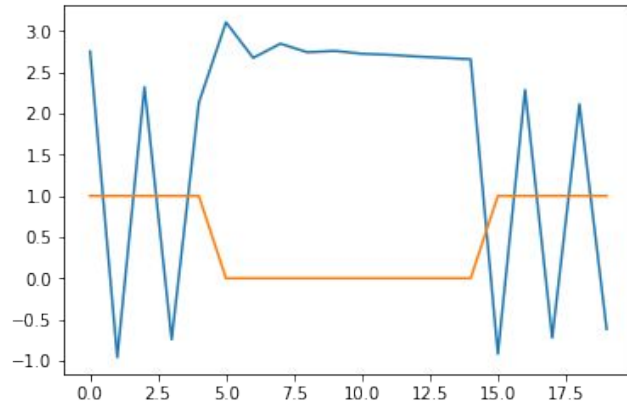
Basic



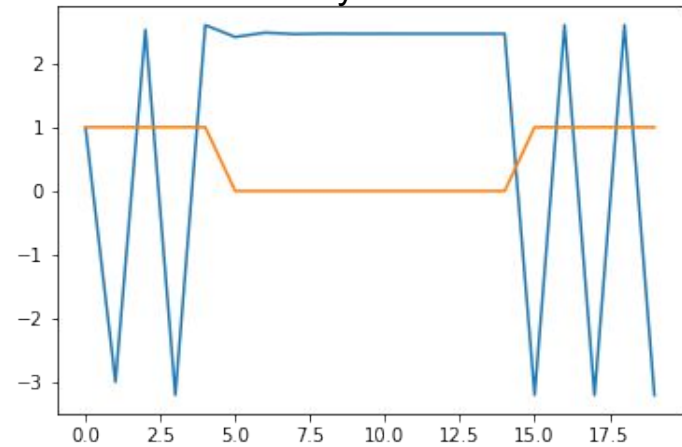
Only F



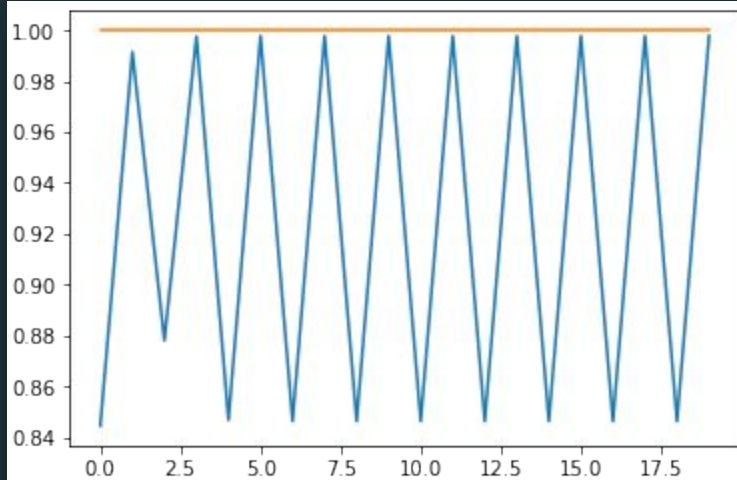
FO



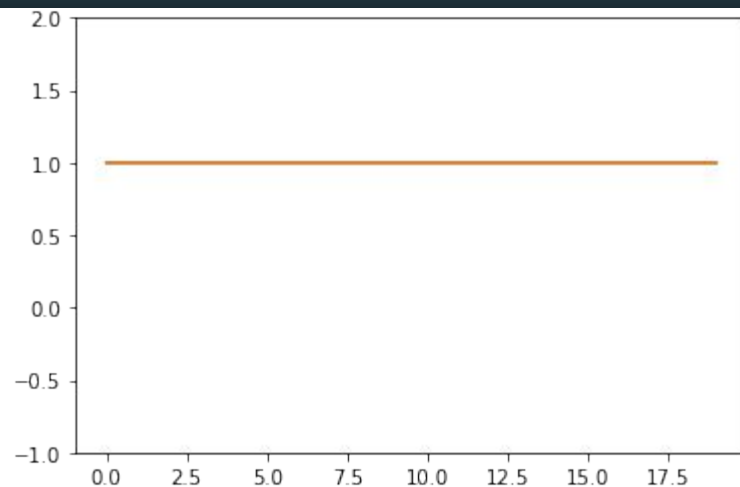
myFO



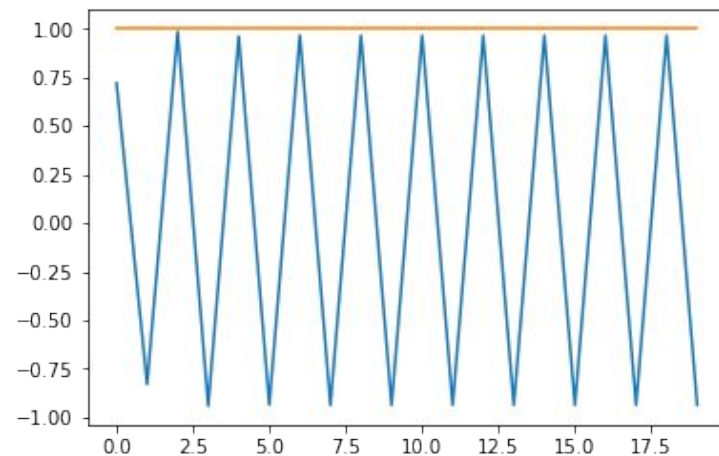
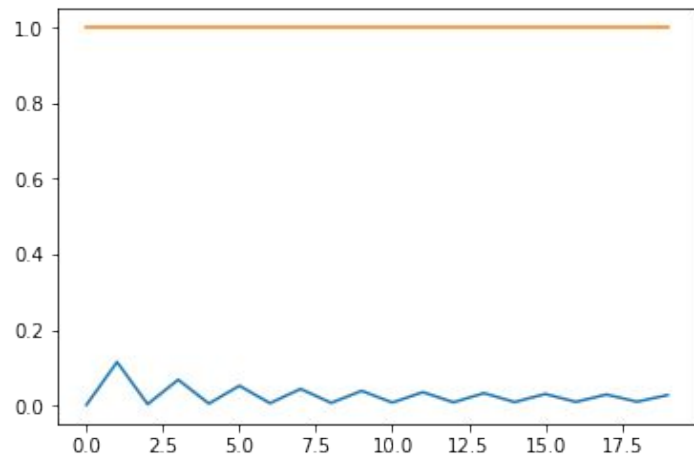




Forget gate  
responses  
myFo - FO



Candidate  
state  
responses  
myFo - FO



# Results

3-HiddenUnit	Basic	F	FO	myFO	IO	O
Accuracy	96.7	91.8	99.89	98.22	80.7	74.47
Iteration Number	267	5942	55	4118	6718	47.419

1-HiddenUnit	Basic	F	FO	myFO	IO	O
Accuracy	81.7	99.85	100	99.93	-	-
Iteration Number	76	67	91	555	-	-

- Hidden size affects generalization dramatically
- Order of Gates important
  - Learning Time
  - Effectiveness of gates
- Less gates can give hard responses

# MNIST

Unit - Accuracy	Basic	F	FO	myFO	IO	O
30	97.22	96.22	96.71	96.96	96.96	95.7
15	93.93	92.67	90.2	95.2	95.45	93.68
1	69.94	20.95	21.21	26.51	24.49	18.68

- Only Forget really outperforms
- Basic is good for starting problem without any prior knowledge
- Forget - Output most important
  - Input gives higher result
  - Importance not free from dataset
  - How you fed input

# Sentiment Analysis

Unit - Accuracy	Basic	F	FO	myFO	IO	O
30	83.61	85.32	83.05	80.25	83.13	63.28

- Complex dataset - Simple Solution
- Basic never the worst
- Forget is most important
- Disabling gates
  - Fine-tuning
  - Regularization

# Conclusion

- Less gate hard decisions
  - working like electronic gates
- Importance depends on
  - Order
  - Dataset
  - Input type
- Start with Basic LSTM
- Gate disabling for regularization