

# RNN BoolQ Classification

Yelin Zhang

# Preprocessing

- Glove Wikipedia Embeddings
  - Lowercase
  - Remove punctuation and non ascii characters
- Truncating & Padding
  - maximum question length 21
  - 99% maximum passage length 275
- Concatenation of question and passage
  - 8x vectors with 0 as separator

# Input/Output

- After Preprocessing: 304 x 300
- Dimensional reduction for input
- Input: Batch size x 91800
- Hidden: Hidden size x Hidden size (defined by experiment)
- Output: Batch size x 1

# Model

- 2 GRU layers (Input, Hidden)
- Linear layer (Hidden, Hidden)
- ReLu
- Linear layer (Hidden, 1)
- Sigmoid

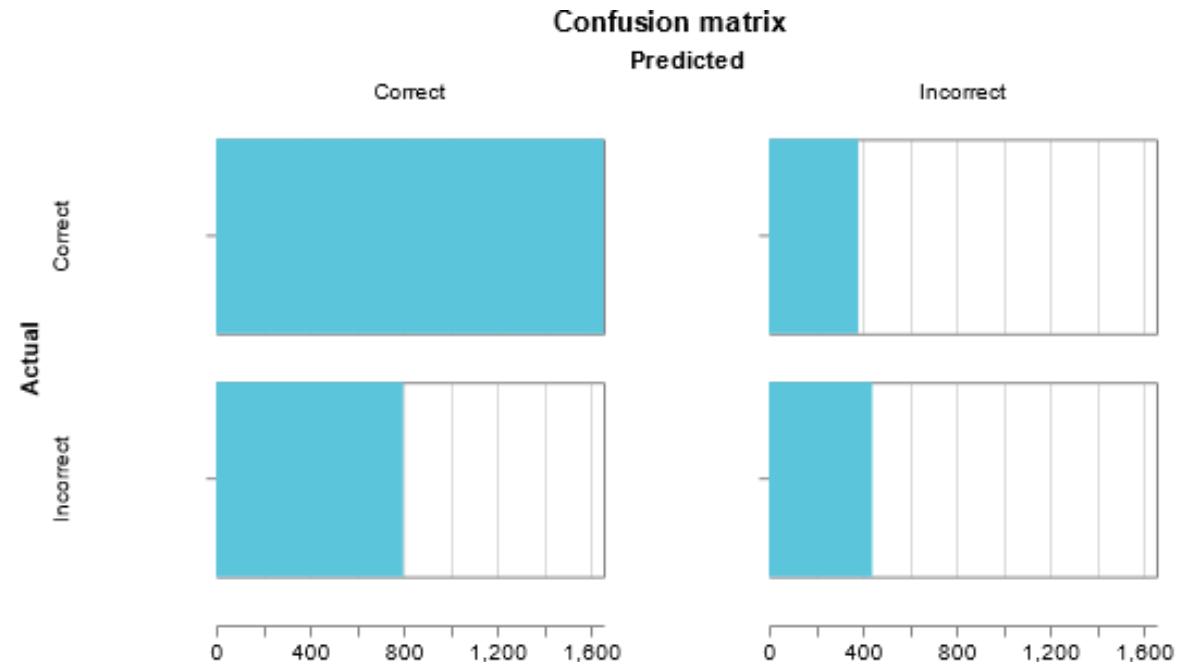
# Experiments

- Hidden size (128,256,512)
- Learning rate (0.01,0.001,0.0001)
- Dropout (0,0.1,0.2)
- Weight decay (0,0.1,0.01)

= 81 experiments

# Results

- Experiments
  - Hidden size does not have a large impact
  - High Learning rate dominates
  - Weight decay 0 dominates
  - Dropout 0 dominates
- Best result
  - 256, 0.001, 0, 0
  - Balanced accuracy:
    - Train 83%, Validation 55%, Test 54%
  - Accuracy 64%



# Interpretation

- Not that good input -> Variable hidden size
- 50 epochs -> Large learning rate
- Few epochs -> Weight decay 0
- Few epochs -> Dropout 0

# Learnings

- Too many checkpoints
- Wandb sweeps are good
- Wandb confusion matrix is not easily createable
- Pytorch Lightning documentation is sometimes cryptic
- Should have tested LSTM vs GRU