

# Stanford Sentiment Treebank with 5 labels (SST-5)

Vinicius Ambrosi, Gilyoung Cheong, and Dohoon Kim

# Motivation

Sentiment Analysis

Q. Can we train a model that can rate an online comment from 1 to 5 stars?

Constraint: small dataset (< 10k)

---

# What is SST-5?

## Term

- SST-5 means Stanford Sentiment Treebank with 5 labels.
- It is a dataset utilized for sentiment analysis.

## Details

- 11,855 sentences sourced from movie reviews, labeled by three human judges and into 5 ratings.
- 5 rating problems are considered VERY HARD.

## State-of-Art

- SOTA: test accuracy of **59.8**
- Top 5: **55.5**
- Rank 5 only used BERT + dropout

# Reaching for Top 5 with a Sentence Transformer

## What is a sentence transformer?

- A sentence transformer is essentially a transformer (e.g., BERT) followed by a contrastive loss function, which helps us retrain the vectors so that they remember which sentences are close to each other.
- Sentence  $\rightarrow$  Vector
- We use the following pre-trained sentence transformers and their fine-tuned models: General Text Embeddings (GTE), Stella

# Hyperparameter Tuning (FNN on GTE)

## Hyperparameters

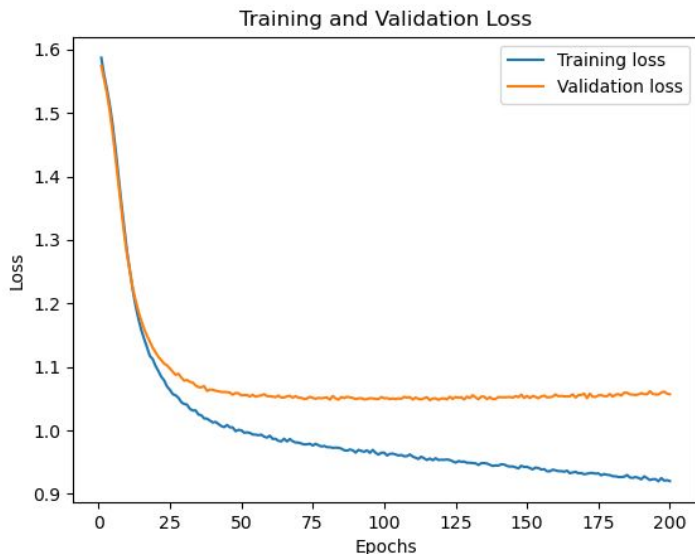
- Optimizer
- Learning rate
- Batch size
- Number of layers and neurons
- Regularization and dropout

## Tools

- Ray Tune
- ASHA
- Optuna

# Hyperparameter Tuning

## Conclusions



- Common observations
  - Lower learning rate
  - Shallow network with small number of neurons
  - Dropout improves performance
- Common issue: overfitting
- Possible solutions:
  - Feature reduction
  - Data augmentation
  - Transformer fine-tuning

# Fine-Tuning Process

## 1. Undersample

### Example

- # of 1-star ratings: 1092
- # of 2-star ratings: 2218
- # of 3-star ratings: 1624
- # of 4-star ratings: 2322
- # of 5-star ratings: 1288

We can random undersample all the ratings but 1-star to have 1092 samples

## 2. A, P, N triple

### Example

Anchor (5): "Good fun, good action, good acting, good dialogue, good pace, good cinematography."

Positive (5): "Allen 's funniest and most likeable movie in years ."

Negative (1): "A real snooze."

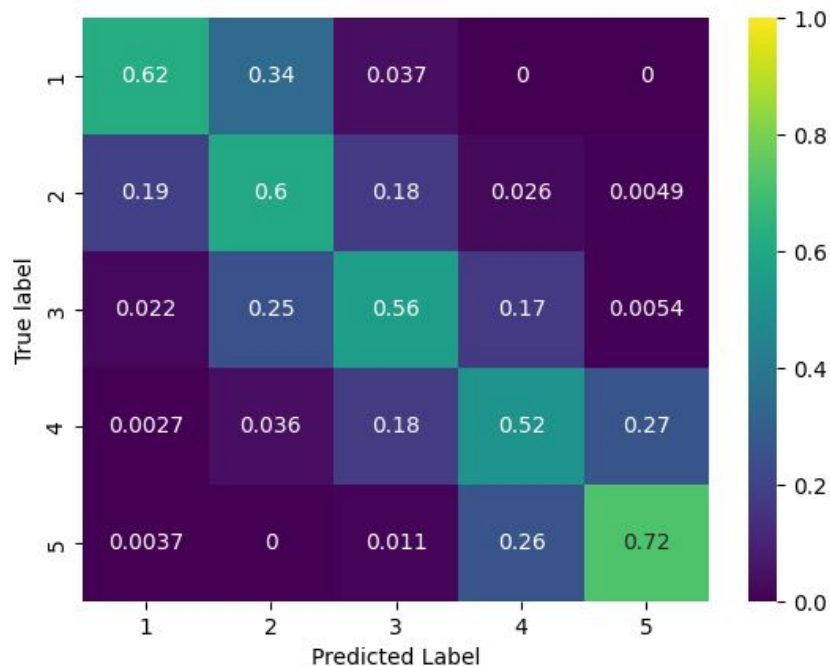
## 3. Freeze layers

### Transformer

We found that it is helpful to freeze transformer layers and only unfreeze (part or whole) dense layers.

# Conclusion

- With a fine-tuned lightweight sentence transformer and a shallow neural network, we achieved **56.9%** accuracy (better than **4th** on leaderboard).
- With a heavier sentence transformer and fine-tuning, we achieved **58.6%** accuracy (better than **3rd** on the leaderboard).





# Future Directions

- Ordinal regression
  - Feature reduction
  - Data augmentation
  - Different fine-tuning (e.g., triples and loss functions)
-



Thank you!