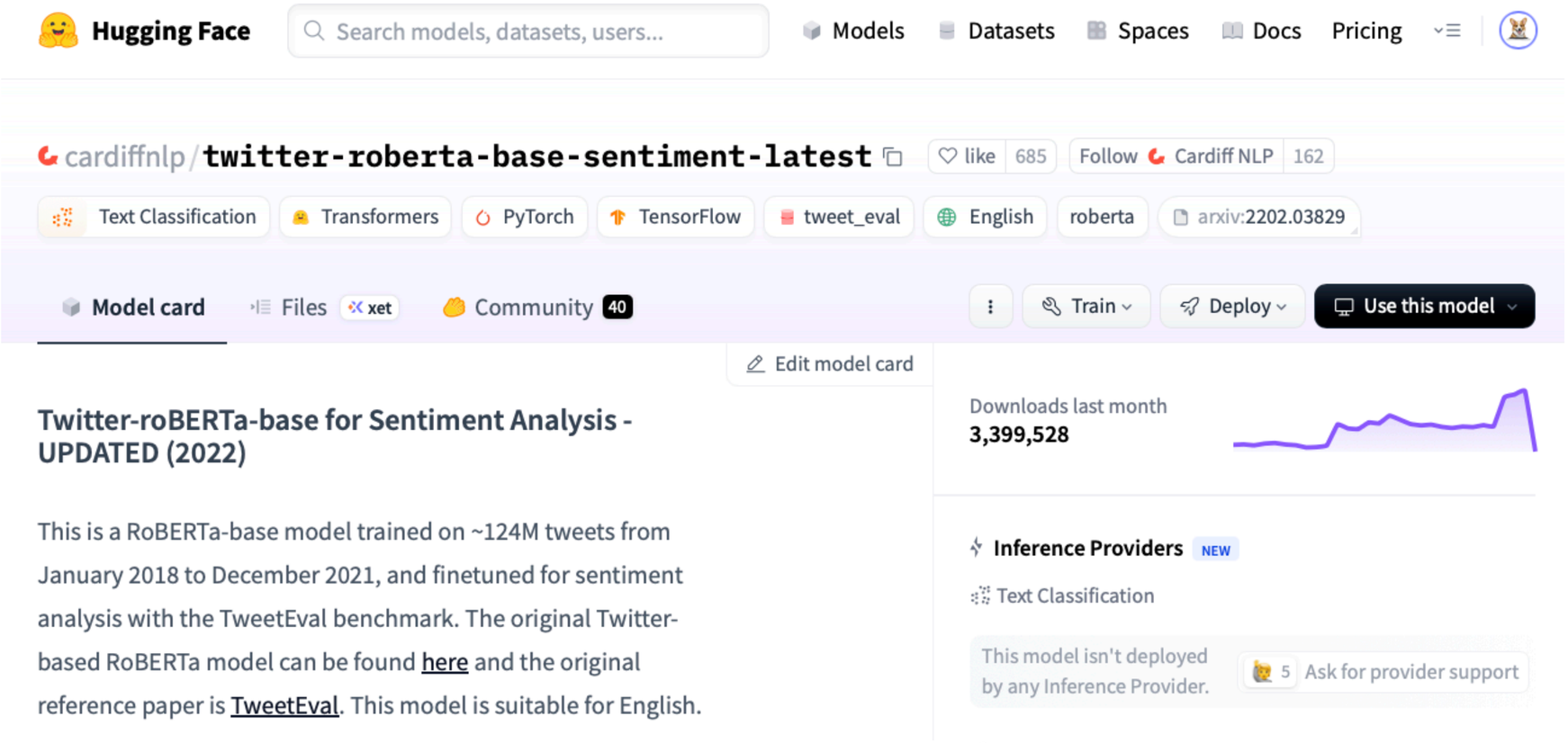
Assignment 4

2025 딥러닝

박재욱 / 융합소프트웨어학부

Robustly optimized BERT

- RoBERTa : mainly just train BERT for longer and remove next sentence prediction!
- twitter 기반 모델 - 구어체, 이모지, 해시태그 잘 처리

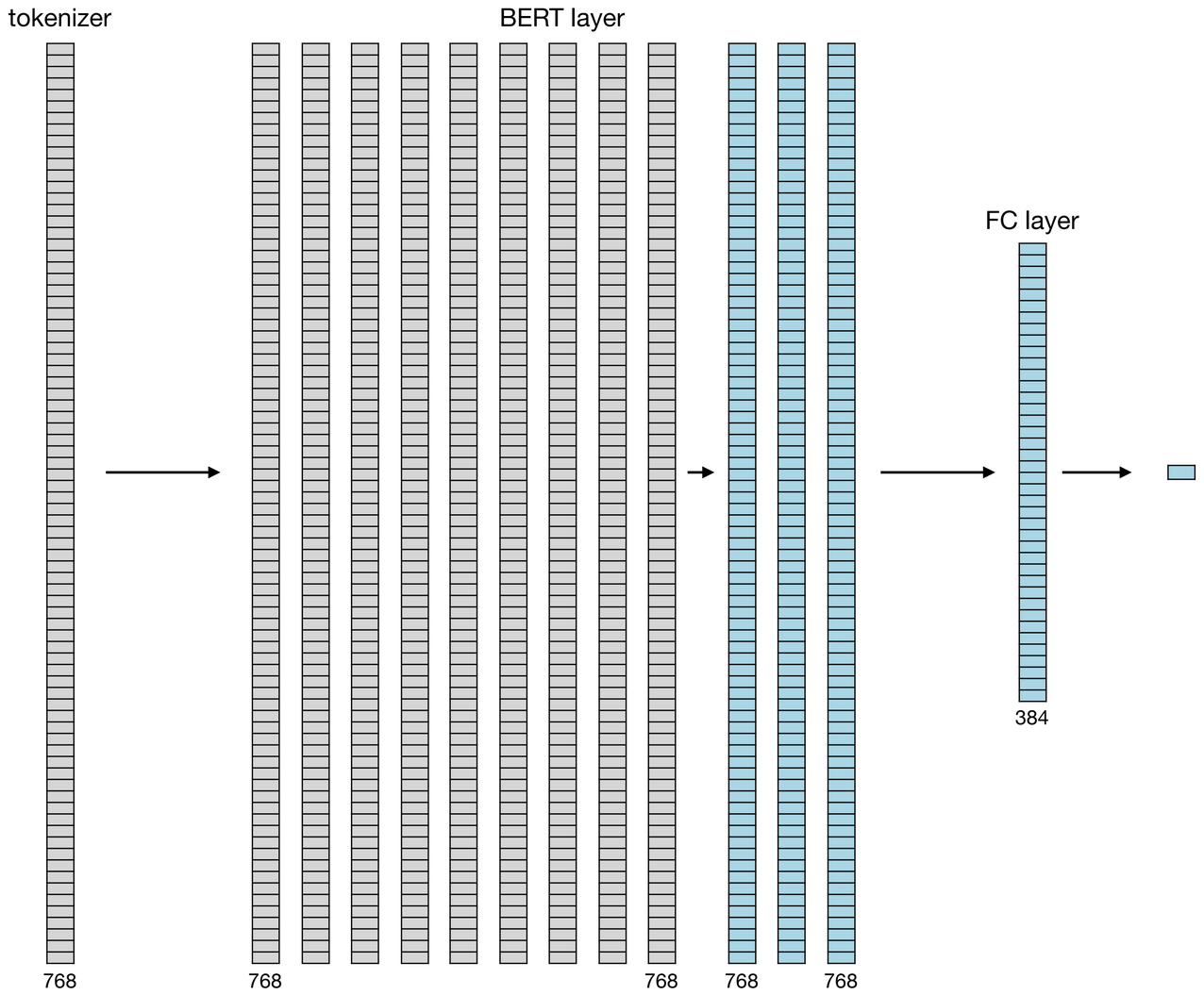


모델 설명

- twitter-roberta 토크나이저 사용
- 입력 토큰 최대 길이 512
- batch_size 64
- twitter-roberta 9개 layer freeze, 3개 layer만 학습
- GradScaler + autocast = 16bit floating point
- Learning rate

  RoBERTa layer = 2e-5

  FC layer = 1e-4
- weight_decay=0.01
- 최고 기록 갱신할 때마다 weight 저장



학습 결과

- 최고 기록을 갱신할 때마다 모델의 weight 저장
- test_acc 4번 연속 오르지 않으면 종료하고 best weight 로드

```
Training: 100%|██████████| 391/391 [02:30<00:00,  2.60it/s, loss=0.12, train_acc=0.957]
Epoch 6 summary - avg loss: 0.1197, Train acc: 0.9572, Test acc: 0.9437

Training: 100%|██████████| 391/391 [02:30<00:00,  2.60it/s, loss=0.104, train_acc=0.963]
Epoch 7 summary - avg loss: 0.1044, Train acc: 0.9631, Test acc: 0.9373

Training: 100%|██████████| 391/391 [02:30<00:00,  2.61it/s, loss=0.0937, train_acc=0.967]
Epoch 8 summary - avg loss: 0.0937, Train acc: 0.9668, Test acc: 0.9444

Training: 100%|██████████| 391/391 [02:30<00:00,  2.60it/s, loss=0.083, train_acc=0.97]
Epoch 9 summary - avg loss: 0.0830, Train acc: 0.9698, Test acc: 0.9431

Training: 100%|██████████| 391/391 [02:30<00:00,  2.60it/s, loss=0.0681, train_acc=0.976]
Epoch 10 summary - avg loss: 0.0681, Train acc: 0.9756, Test acc: 0.9420

Training: 100%|██████████| 391/391 [02:30<00:00,  2.60it/s, loss=0.06, train_acc=0.979]
Epoch 11 summary - avg loss: 0.0600, Train acc: 0.9790, Test acc: 0.9413

Training: 100%|██████████| 391/391 [02:30<00:00,  2.60it/s, loss=0.0528, train_acc=0.982]
Epoch 12 summary - avg loss: 0.0528, Train acc: 0.9817, Test acc: 0.9398


Early stopping.

Best test acc: 0.9444
Applying optimal model weights.
```