Edit Share

Sentiment-Classification with Bert-pretrained

With Bert-pretrained model, to make the highest accuracy in 'yelp' restaurant data, modify variable variance (e.g batch_size, learning rate, learning rate scheduler, weight_decay)

youngjin yoon

standard-setting: stable acc, stable declination of loss except val loss at step 50

with lr-schedular1: the highest acc, after third epoch a little rise of val loss -> plan to make lr change per epoch more smoothly(0.65 -> 0.95)

with lr-schedular2: relatively high loss and low acc until the mid point of training, after modifying lr at third epoch (lr*0.65**2) -> can see meaningful declination of loss and rise of acc => appropriate lr is important to make lower loss

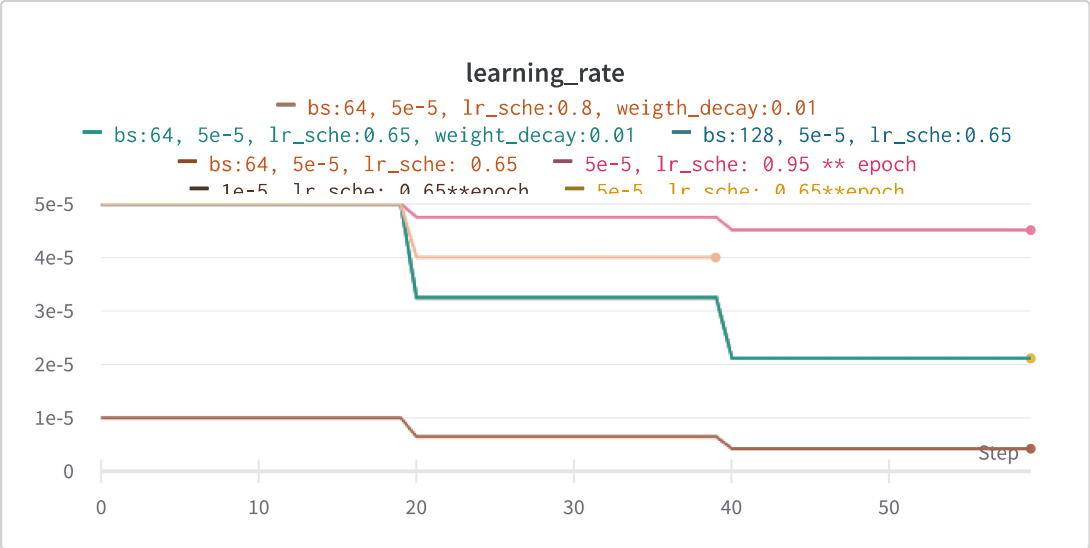
with lr-scheduler3: for the pupose of declining loss after third epoch, make lr scheduler more smoothly but couldn't find any meaningfulness

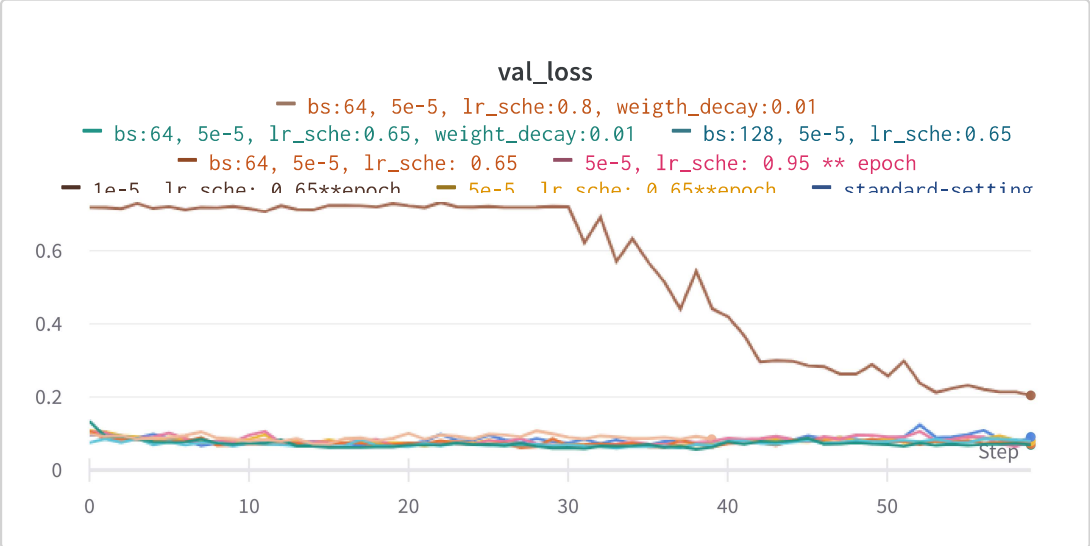
make batch_size bigger for using GPU efficiently (32->64) -> show meaningful rise of acc

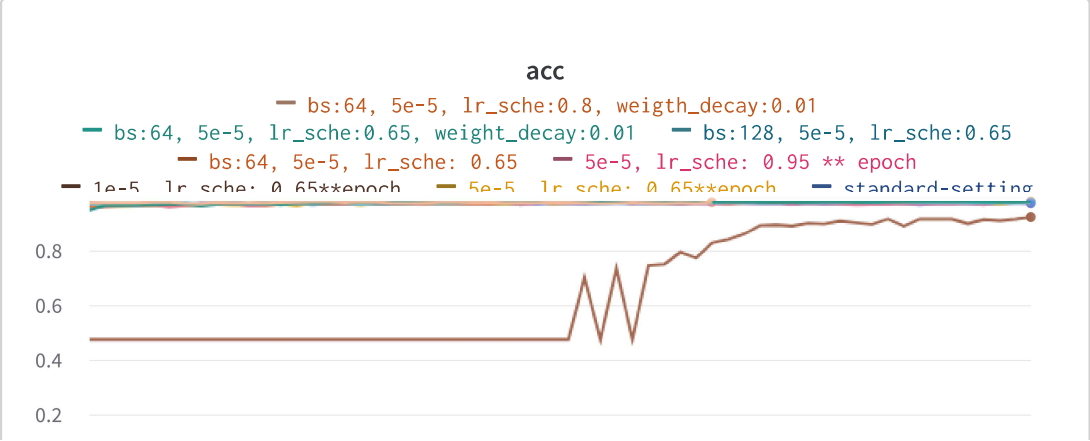
make batch_size "more" bigger (64->128) -> in third epoch there is overfitting problem

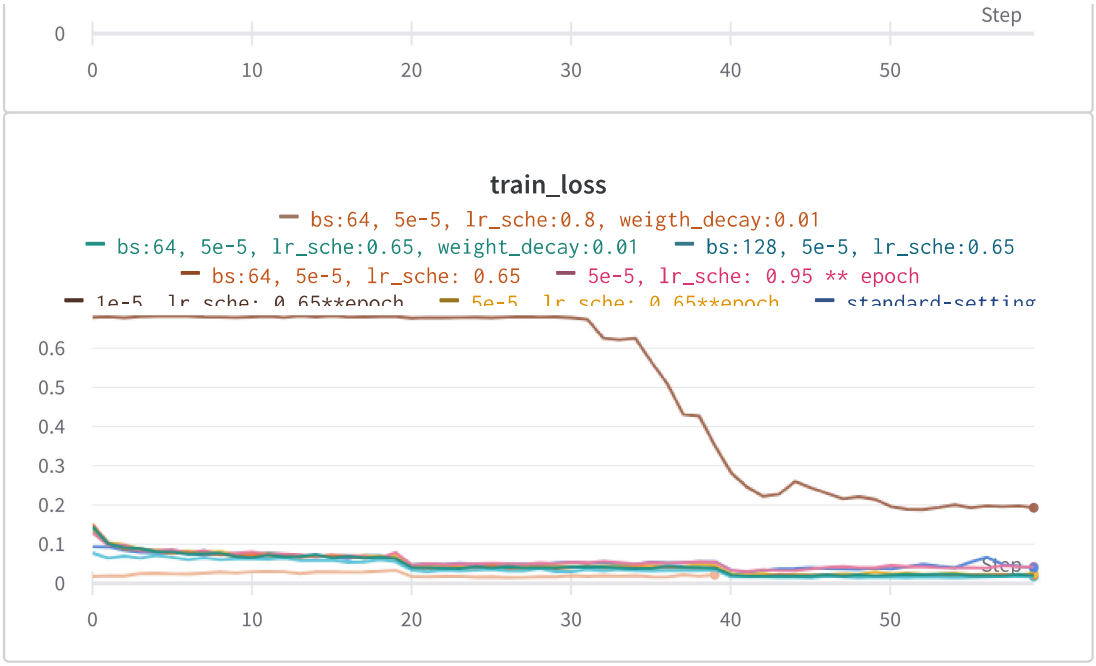
so fix the batch_size 32 and use weight_decay 0.01 for preventing from overfitting problem

Section 1


learning_rate
bs:64, 5e-5, lr_sche:0.8, weigth_decay:0.01
bs:64, 5e-5, lr_sche:0.65, weight_decay:0.01
bs:128, 5e-5, lr_sche:0.65
bs:64, 5e-5, lr_sche: 0.65
5e-5, lr_sche: 0.95 ** epoch
5e-5
4e-5
3e-5
2e-5
1e-5
0
Step
0
10
20
30
40
50


val_loss
bs:64, 5e-5, lr_sche:0.8, weigth_decay:0.01
bs:64, 5e-5, lr_sche:0.65, weight_decay:0.01
bs:128, 5e-5, lr_sche:0.65
bs:64, 5e-5, lr_sche: 0.65
5e-5, lr_sche: 0.95 ** epoch
standard-setting
0.6
0.4
0.2
0
Step
0
10
20
30
40
50


acc
bs:64, 5e-5, lr_sche:0.8, weigth_decay:0.01
bs:64, 5e-5, lr_sche:0.65, weight_decay:0.01
bs:128, 5e-5, lr_sche:0.65
bs:64, 5e-5, lr_sche: 0.65
5e-5, lr_sche: 0.95 ** epoch
standard-setting
0.8
0.6
0.4
0.2



Created with ♥ on Weights & Biases.

https://wandb.ai/team_koowater/Bert/reports/Sentiment-Classification-with-Bert-pretrained--VmlldzoyMzMxODA5