

Report HW5

Giorgio Giannone

July 30, 2018

Introduction

Our goal is to implement a Recurrent Network to classify sentences.

Result

We report the results for three different settings in Fig.1, Fig.2, Fig.3.

```
Created model with fresh parameters.
epoch 0 learning rate 0.0050 epoch-time 32.9012 loss 2.11097897 accuracy [0.30770131]
        dev_set, loss 1.60877854, accuracy [0.33333333]
epoch 1 learning rate 0.0050 epoch-time 31.3758 loss 1.63456667 accuracy [0.36821161]
        dev_set, loss 1.47367989, accuracy [0.39237057]
epoch 2 learning rate 0.0050 epoch-time 31.5113 loss 1.50977801 accuracy [0.39466292]
        dev_set, loss 1.56196954, accuracy [0.36148955]
epoch 3 learning rate 0.0050 epoch-time 31.2687 loss 1.43476861 accuracy [0.41022940]
        dev_set, loss 1.40771552, accuracy [0.40962761]
epoch 4 learning rate 0.0050 epoch-time 31.5710 loss 1.38088751 accuracy [0.42638109]
        dev_set, loss 1.52418794, accuracy [0.39237057]
epoch 5 learning rate 0.0050 epoch-time 31.4180 loss 1.34120978 accuracy [0.45224719]
        dev_set, loss 1.46210134, accuracy [0.38419619]
epoch 6 learning rate 0.0050 epoch-time 31.4916 loss 1.30470692 accuracy [0.45798221]
        dev_set, loss 1.46024762, accuracy [0.38782925]
epoch 7 learning rate 0.0050 epoch-time 31.3504 loss 1.28670949 accuracy [0.46301498]
        dev_set, loss 1.39735005, accuracy [0.37511353]
epoch 8 learning rate 0.0050 epoch-time 31.2386 loss 1.25845150 accuracy [0.48384831]
        dev_set, loss 1.42106053, accuracy [0.41326067]
epoch 9 learning rate 0.0050 epoch-time 32.0488 loss 1.24016763 accuracy [0.48747659]
        dev_set, loss 1.43561046, accuracy [0.39237057]
epoch 10 learning rate 0.0050 epoch-time 32.1171 loss 1.20891345 accuracy [0.50152154]
        dev_set, loss 1.41293253, accuracy [0.40599455]
epoch 11 learning rate 0.0050 epoch-time 31.8050 loss 1.18300931 accuracy [0.50690543]
        dev_set, loss 1.50314155, accuracy [0.38873751]
epoch 12 learning rate 0.0050 epoch-time 32.1189 loss 1.16487882 accuracy [0.52083333]
        dev_set, loss 1.41869863, accuracy [0.38328792]
epoch 13 learning rate 0.0050 epoch-time 32.4826 loss 1.13877087 accuracy [0.53464419]
        dev_set, loss 1.62137580, accuracy [0.38964578]
epoch 14 learning rate 0.0050 epoch-time 32.0792 loss 1.11523402 accuracy [0.54705056]
        dev_set, loss 1.43020480, accuracy [0.39237057]
epoch 15 learning rate 0.0050 epoch-time 32.3334 loss 1.08204058 accuracy [0.55933989]
        dev_set, loss 1.44700261, accuracy [0.40962761]
epoch 16 learning rate 0.0050 epoch-time 31.3303 loss 1.05155138 accuracy [0.58438670]
        dev_set, loss 1.46970465, accuracy [0.38692098]
epoch 17 learning rate 0.0050 epoch-time 31.2217 loss 1.02116211 accuracy [0.58906835]
        dev_set, loss 1.49595737, accuracy [0.39509537]
epoch 18 learning rate 0.0050 epoch-time 32.2241 loss 0.98521214 accuracy [0.60779494]
        dev_set, loss 1.50829836, accuracy [0.38147139]
epoch 19 learning rate 0.0050 epoch-time 31.7273 loss 0.94095709 accuracy [0.62862828]
        dev_set, loss 1.55700331, accuracy [0.39509537]
```

Figure 1: result using a vanilla RNN cell

Experiments

To obtain the result with the LSTM cell with the following modifications:

- Optimizer using Gradient Descent with momentum 0.9
- More penalization (0.001) for the regularization of the loss function
- Number neurons (1024)

```

Reading model parameters from ./train
epoch 0 learning rate 0.0050 epoch-time 51.0500 loss 1.38754988 accuracy [0.42532772]
        dev_set, loss 1.42388931, accuracy [0.38056312]
epoch 1 learning rate 0.0050 epoch-time 48.6760 loss 1.37138678 accuracy [0.43867041]
        dev_set, loss 1.39941585, accuracy [0.40781108]
epoch 2 learning rate 0.0050 epoch-time 47.9423 loss 1.35809335 accuracy [0.43609551]
        dev_set, loss 1.46914084, accuracy [0.36875568]
epoch 3 learning rate 0.0050 epoch-time 48.0108 loss 1.34756098 accuracy [0.43153090]
        dev_set, loss 1.36732529, accuracy [0.41961853]
epoch 4 learning rate 0.0050 epoch-time 50.1313 loss 1.33527090 accuracy [0.43691479]
        dev_set, loss 1.58202732, accuracy [0.37965486]
epoch 5 learning rate 0.0050 epoch-time 50.0181 loss 1.32059689 accuracy [0.44850187]
        dev_set, loss 1.42745709, accuracy [0.40417802]
epoch 6 learning rate 0.0050 epoch-time 48.1010 loss 1.30889795 accuracy [0.44803371]
        dev_set, loss 1.49878353, accuracy [0.38419619]
epoch 7 learning rate 0.0050 epoch-time 48.0050 loss 1.30604999 accuracy [0.45049157]
        dev_set, loss 1.39994640, accuracy [0.39418710]
epoch 8 learning rate 0.0050 epoch-time 47.2951 loss 1.29630032 accuracy [0.46090824]
        dev_set, loss 1.42852893, accuracy [0.39691190]
epoch 9 learning rate 0.0050 epoch-time 47.9268 loss 1.29290217 accuracy [0.45529026]
        dev_set, loss 1.39950530, accuracy [0.39963669]
epoch 10 learning rate 0.0050 epoch-time 47.4622 loss 1.28084506 accuracy [0.46371723]
        dev_set, loss 1.39580597, accuracy [0.41235241]
epoch 11 learning rate 0.0050 epoch-time 47.2019 loss 1.26346182 accuracy [0.46477060]
        dev_set, loss 1.40809890, accuracy [0.39418710]
epoch 12 learning rate 0.0050 epoch-time 47.6175 loss 1.26520960 accuracy [0.46839888]
        dev_set, loss 1.35865332, accuracy [0.37965486]
epoch 13 learning rate 0.0050 epoch-time 47.4070 loss 1.25726568 accuracy [0.47483614]
        dev_set, loss 1.50123355, accuracy [0.40508629]
epoch 14 learning rate 0.0050 epoch-time 47.2803 loss 1.25062401 accuracy [0.47038858]
        dev_set, loss 1.37384695, accuracy [0.41235241]
epoch 15 learning rate 0.0050 epoch-time 47.3318 loss 1.23990349 accuracy [0.47109082]
        dev_set, loss 1.37607190, accuracy [0.41871026]
epoch 16 learning rate 0.0050 epoch-time 47.3523 loss 1.23672519 accuracy [0.47881554]
        dev_set, loss 1.39094461, accuracy [0.38510445]
epoch 17 learning rate 0.0050 epoch-time 47.1053 loss 1.23233960 accuracy [0.47869850]
        dev_set, loss 1.42942871, accuracy [0.39691190]
epoch 18 learning rate 0.0050 epoch-time 47.0058 loss 1.22541508 accuracy [0.48712547]
        dev_set, loss 1.38443186, accuracy [0.37783833]
epoch 19 learning rate 0.0050 epoch-time 47.7169 loss 1.21852418 accuracy [0.49016854]
        dev_set, loss 1.36496251, accuracy [0.41053588]

```

Figure 2: result using a GRU cell

- Batch size 256

I also tried to increase the number of layers, but it didn't improve the dev accuracy.

```
Created model with fresh parameters.
epoch 0 learning rate 0.0050 epoch-time 8.7289 loss 1.66815835 accuracy [0.29377341]
        dev_set, loss 1.54546450, accuracy [0.36966394]
epoch 1 learning rate 0.0050 epoch-time 7.5723 loss 1.45096280 accuracy [0.38085206]
        dev_set, loss 1.50233146, accuracy [0.32515895]
epoch 2 learning rate 0.0050 epoch-time 7.5753 loss 1.39380307 accuracy [0.40648408]
        dev_set, loss 1.36994139, accuracy [0.40690282]
epoch 3 learning rate 0.0050 epoch-time 7.5290 loss 1.36990136 accuracy [0.41081461]
        dev_set, loss 1.34478316, accuracy [0.39963669]
epoch 4 learning rate 0.0050 epoch-time 7.4939 loss 1.35660210 accuracy [0.42731742]
        dev_set, loss 1.45197448, accuracy [0.34695731]
epoch 5 learning rate 0.0050 epoch-time 7.5285 loss 1.31613368 accuracy [0.44171348]
        dev_set, loss 1.38763465, accuracy [0.40417802]
epoch 6 learning rate 0.0050 epoch-time 7.5211 loss 1.31512056 accuracy [0.44323502]
        dev_set, loss 1.41767427, accuracy [0.36693915]
epoch 7 learning rate 0.0050 epoch-time 7.5272 loss 1.29512843 accuracy [0.44077715]
        dev_set, loss 1.37238883, accuracy [0.42960945]
epoch 8 learning rate 0.0050 epoch-time 7.4507 loss 1.28180716 accuracy [0.45294944]
        dev_set, loss 1.39305066, accuracy [0.38964578]
epoch 9 learning rate 0.0050 epoch-time 7.4735 loss 1.29044320 accuracy [0.44990637]
        dev_set, loss 1.35459072, accuracy [0.41961853]
epoch 10 learning rate 0.0050 epoch-time 7.4759 loss 1.29671520 accuracy [0.45259831]
        dev_set, loss 1.40623575, accuracy [0.38328792]
epoch 11 learning rate 0.0050 epoch-time 7.4674 loss 1.29629902 accuracy [0.44651217]
        dev_set, loss 1.57993164, accuracy [0.36330609]
epoch 12 learning rate 0.0050 epoch-time 7.3945 loss 1.28447400 accuracy [0.46067416]
        dev_set, loss 1.38456855, accuracy [0.42325159]
epoch 13 learning rate 0.0050 epoch-time 7.5060 loss 1.25951949 accuracy [0.46371723]
        dev_set, loss 1.33905774, accuracy [0.42597639]
epoch 14 learning rate 0.0050 epoch-time 7.4896 loss 1.25539447 accuracy [0.46313202]
        dev_set, loss 1.37166449, accuracy [0.41780200]
epoch 15 learning rate 0.0050 epoch-time 7.4855 loss 1.24867537 accuracy [0.46886704]
        dev_set, loss 1.40619815, accuracy [0.41507720]
epoch 16 learning rate 0.0050 epoch-time 7.5077 loss 1.32138309 accuracy [0.44990637]
        dev_set, loss 1.36746850, accuracy [0.41416894]
epoch 17 learning rate 0.0050 epoch-time 7.4240 loss 1.24581615 accuracy [0.47460206]
        dev_set, loss 1.39031697, accuracy [0.40781108]
epoch 18 learning rate 0.0050 epoch-time 7.5159 loss 1.24844862 accuracy [0.47471910]
        dev_set, loss 1.38235298, accuracy [0.41598547]
epoch 19 learning rate 0.0050 epoch-time 7.5292 loss 1.23680881 accuracy [0.47717697]
        dev_set, loss 1.34171938, accuracy [0.43415077]
```

Figure 3: result using a LSTM cell