Headlines Generation

Assignment 5

Luigi Quarantiello

Code snippets

```
#Snippet for scraping the latest headlines from Lercio.it
base_URL = 'https://www.lercio.it/category/brevi/page/'

for page_count in range(1,73):
    page = requests.get(base_URL + str(page_count))
    soup = BeautifulSoup(page.content, 'html.parser')

    for div in soup.find_all('div', {'class': 'jeg_sidebar'}):
        div.decompose()

    articles = soup.find_all('article')

    with open('headlines/titles_ultima.csv', 'a') as file:
        writer = csv.writer(file, delimiter=',')

        # get the title from the two highlighted news only once
        if page_count == 1:
            writer.writerow([articles[0].find('h2').a.text])
            writer.writerow([articles[1].find('h2').a.text])

        for element in articles[2:]:
            title = element.find('h3')
            headline = title.a.text
            writer.writerow([headline])
```

```
#Implementation used
https://github.com/crazydonkey200/tensorflow-char-rnn
```

```
#Parameters grid for model selection
{
    "hidden_size" : [128,256,512],
    "num_layers" : [2,3],
    "dropout" : [0.5],
    "batch_size" : [32,64,128]
}
```

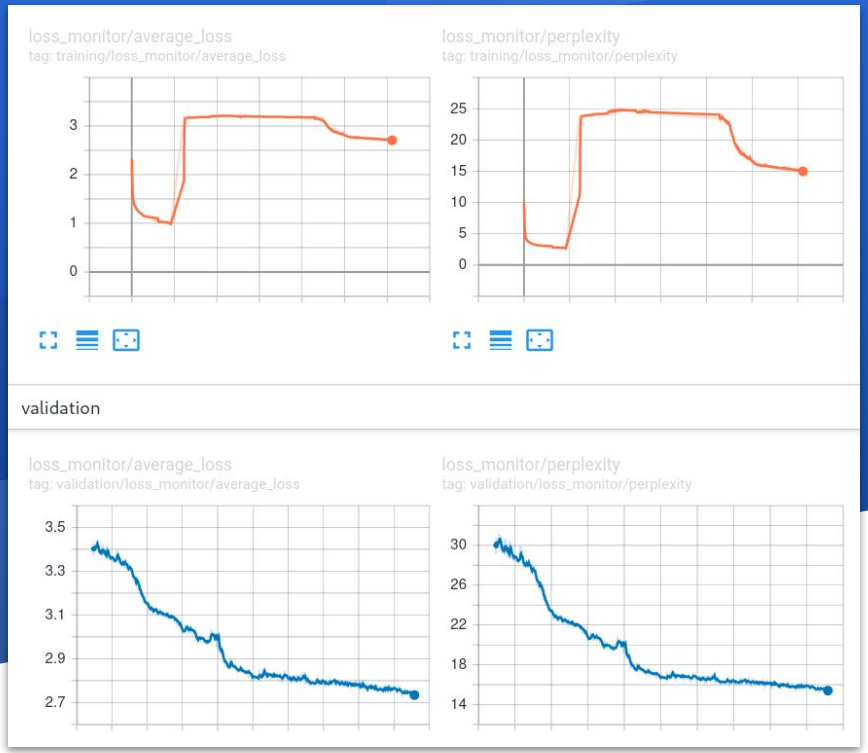
```
!python train.py --data_file=dataset.txt --hidden_size=hs
                 --num_layers=nl --model='lstm'
                 --dropout=d --batch_size=bs
```

```
#Generation of new headlines
!python sample.py --init_dir=output --start_text=st
                  --length=l
```

Results - Tensorboard

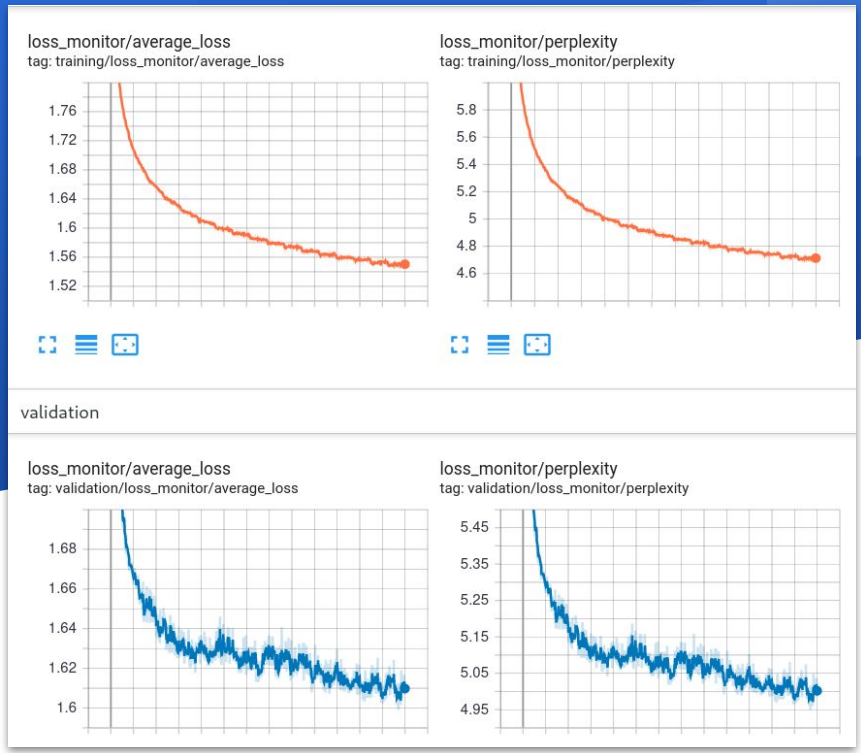
```
{
    "hidden_size" : [512],  "num_layers" : [2],
    "dropout" : [0.5],  "batch_size" : [32]
}
```

"best_valid_ppl": 4.39167




```
{
    "hidden_size" : [128],  "num_layers" : [3],
    "dropout" : [0.5],  "batch_size" : [32]
}
```

"best_valid_ppl": 4.94897,



Results - Samples


```
{
    "hidden_size" : [512],  "num_layers" : [2],
    "dropout" : [0.5],  "batch_size" : [32]
}
```

```
"best_valid_ppl": 4.39167
```


- Sciopero M5S, riscaldamento da selfie su Fratello
- Scienza: spunta la Nazionale alla strada per fare cervello per la nazionale
- Di Maio rimente
- Falsi alla lega ma nessuno se ne accorge
- Negoziamento ottiene la Madonna


```
{
    "hidden_size" : [128],  "num_layers" : [3],
    "dropout" : [0.5],  "batch_size" : [32]
}
```

```
"best_valid_ppl": 4.94897,
```


- Di Maio: “La mectaliziane lascia al Lattrino Bozzero”
- Cina, 2006 anti-sondaggi lunda torna
- Romani fa direttori sospesi

Final considerations

- **The model with the lowest perplexity seems to have learned the topics and the style of the Lercio newspage**

- **It may need a deeper model selection, possibly using a bigger dataset**
- **The training is really computationally expensive**