A3-write-up

Task 1-3 Evaluation and extended experiment

a) Report perplexity on training set, as the dataset is very small.

I trian the RNNLanguageModel for five epochs and the perplexity results are the followings.

```
Epoch 1, Perplexity: 6.993050575256348
Epoch 2, Perplexity: 1.0415560007095337
Epoch 3, Perplexity: 1.0183625221252441
Epoch 4, Perplexity: 1.0150909423828125
Epoch 5, Perplexity: 1.0140013694763184
```

b) Generate some sentences; quality is not graded.

I set the `max_length = 50` and set `temperature = 2` to add some randoms. I generate five sentences and the results are the followings.

```
Generated Sentence Indices: 子 曰 ： 咎 系 裘 生 巽 占 沮 便 荆 肥 木 他 腥 畜 苟 王 尧 祝 沮 图 乘 宾 佾 割 废 而 虽 诺 衣 口 封 莞 喟 诗 助 材 念 罔 损 道 宫 匡 质 桴 亡 礼 羊 列 鞮 善
Generated Sentence Indices: 子 曰 ： 吉 享 冕 巽 给 锦 产 一 日 虐 虐 章 虐 愆 羹 请 母 由 久 。
Generated Sentence Indices: 子 曰 ： 同 撤 证 踖 德 廉 忿 其 黄 玄 穆 本 挚 报 斯 釜 惜 壮 止 冯 问 乎 雌 狷 区 别 彬 棘 士 赍 尝 悲 余 贾 童 浸 且 予 细 族 肸 洁 鞮 , 袍 阼 亚 兕 独 各
Generated Sentence Indices: 子 曰 ： 姓 虎 星 诱 侗 侗 舜 侮 为 修 宫 陶 饰 吉 探 均 施 则 弓 悾 亵 使 争 俭 利 绀 弓 用 兢 鲍 劝 孟 简 拖 崩 圃 费 色 颠 厩 民 诱 空 优 诔 火 观 悾 牡 讷
Generated Sentence Indices: 子 曰 ： 首 忿 柴 矜 谋 物 澹 门 盍 他 诔 阈 躁 伊 馈 衣 迹 愆 洫 门 意 获 容 虐 虐 反 帝 虐 国 尚 志 甚 对 病 翼 萧 忿 没 目 失 卿 罢 萧 静 虐 希 虐 伐 本 憎
```

c) Compare the perplexity on two conditions: randomly initialized embeddings vs. with

pretrained embeddings (from A2).

This part uses the model form A2 that is SkipGram and I use the embeddings with `num_epochs = 5`, `embedding_dim = 50` and `hidden_dim = 100` . The results are the followings. They are really closed as the dataset is too small.

```
perplexity of pretrained embeddings:
64.01287841796875
perplexity of randomly initialized embeddings:
63.99871826171875
```

Task 2-3 Train, evaluate, and save

a) Use greedy search to obtain the predicted labels on test set, i.e., pick the highest

probability label for each time step.

b) Report F-1 score on test set.

I train for 5 epochs and the results for the validation dataset are the followings.

```
Epoch: 1, Train Loss: 1.3582, Val Loss: 0.9525, Val F1: 0.7367
Epoch: 2, Train Loss: 0.8817, Val Loss: 0.6232, Val F1: 0.7810
Epoch: 3, Train Loss: 0.5739, Val Loss: 0.4090, Val F1: 0.8200
Epoch: 4, Train Loss: 0.3793, Val Loss: 0.2785, Val F1: 0.8299
Epoch: 5, Train Loss: 0.2610, Val Loss: 0.1979, Val F1: 0.8495
```

The results for the test dataset are the followings.

```
Test Loss: 0.21526903327968386
Test F1 Score: 0.8186873603989944
```