

Assignment 3 报告

阳希明 15307130263

Part1:

LSTM 正向传播公式为:

$$\begin{aligned}z &= [h_{t-1}, X_t] \\f_t &= \sigma(W_f * z + b_f) \\i_t &= \sigma(W_i * z + b_i) \\\bar{c}_t &= \tanh(W_c * z + b_c) \\C_t &= f_t * C_{t-1} + i_t * \bar{c}_t \\o_t &= \sigma(W_o * z + b_o) \\h_t &= o_t * \tanh(C_t)\end{aligned}$$

仅考虑每一个式子，有：

$$\begin{aligned}\frac{\partial h_t}{\partial o_t} &= \text{diag}[\tanh(c_t)] \\\frac{\partial h_t}{\partial c_t} &= \text{diag}[o_t * (1 - \tanh(c_t)^2)] \\\frac{\partial c_t}{\partial f_t} &= \text{diag}[c_{t-1}] \\\frac{\partial c_t}{\partial i_t} &= \text{diag}[\bar{c}_t] \\\frac{\partial c_t}{\partial \bar{c}_t} &= \text{diag}[i_t] \\\frac{\partial c_t}{\partial c_{t-1}} &= \text{diag}[f_t]\end{aligned}$$

连接部分式子，有：

$$\begin{aligned}\frac{\partial h_t}{\partial f_t} &= o_t * (1 - \tanh(c_t)^2) * c_{t-1} \\\frac{\partial h_t}{\partial i_t} &= o_t * (1 - \tanh(c_t)^2) * \bar{c}_t \\\frac{\partial h_t}{\partial \bar{c}_t} &= o_t * (1 - \tanh(c_t)^2) * i_t \\\frac{\partial h_t}{\partial c_{t-1}} &= o_t * (1 - \tanh(c_t)^2) * f_t\end{aligned}$$

其中， o_t, f_t, i_t, \bar{c}_t 有：

$$\begin{aligned}o_t &= \sigma(\text{linear}_{o_t}), \text{linear}_{o_t} = W_o * z + b_o \\f_t &= \sigma(\text{linear}_{f_t}), \text{linear}_{f_t} = W_f * z + b_f \\i_t &= \sigma(\text{linear}_{i_t}), \text{linear}_{i_t} = W_i * z + b_i \\\bar{c}_t &= \sigma(\text{linear}_{\bar{c}_t}), \text{linear}_{\bar{c}_t} = W_c * z + b_c\end{aligned}$$

故：

$$\frac{\partial o_t}{\partial \text{linear}_{o_t}} = \text{diag}[o_t * (1 - o_t)]$$

$$\frac{\partial linear_{ot}}{\partial z} = W_o$$

以此类推 4 式，可得：

$$\frac{\partial h_t}{\partial W_o} = \tanh(c_t) * o_t * (1 - o_t) * z$$

$$\frac{\partial h_t}{\partial b_o} = \tanh(c_t) * o_t * (1 - o_t)$$

$$\frac{\partial h_t}{\partial W_f} = o_t * (1 - \tanh(c_t)^2) * c_{t-1} * f_t * (1 - f_t) * z$$

$$\frac{\partial h_t}{\partial b_f} = o_t * (1 - \tanh(c_t)^2) * c_{t-1} * f_t * (1 - f_t)$$

$$\frac{\partial h_t}{\partial W_i} = o_t * (1 - \tanh(c_t)^2) * \bar{c}_t * i_t * (1 - i_t) * z$$

$$\frac{\partial h_t}{\partial b_i} = o_t * (1 - \tanh(c_t)^2) * \bar{c}_t * i_t * (1 - i_t)$$

$$\frac{\partial h_t}{\partial W_c} = o_t * (1 - \tanh(c_t)^2) * i_t * (1 - \bar{c}_t^2) * z$$

$$\frac{\partial h_t}{\partial b_c} = o_t * (1 - \tanh(c_t)^2) * i_t * (1 - \bar{c}_t^2)$$

至此，倒数推导完毕。

对于时序上的传播，就是依照前向传播的相反方向，对于一个 batch 内一定长度的数据，一次计算多次梯度，相加得到总梯度实现沿时间顺序的反响传播。

Part2:

参数初始化:

显然不能用全 0 进行初始化，按照惯常设计，使用随机初始化方法，我使用了 pytorch 的 uniform_函数对几个 linear 层的参数进行初始化。

数据集处理:

使用 43030 首的全唐诗，去除所有标点符号与非中文字符，按照 20 个字进行切片，且步进为 10。对于最后不足 20 个字的片段，不计入最后的数据中

超参数:

Batchsize: 128

Input size: 300

Hidden size: 300

Lr: 0.001

词向量:

使用预先在四库全书上使用 word2vec 方法训练好的字向量作为初始化输入，若没有出现在这些字向量中，则使用随机值。

生成结果(手动断句):

日 暮吟微蒨，嚟娜偃語它。湔謏瞞晏滂，楠燿椅魑熨。
红 颜炽队楼，救婀吁阙姹。袂道赌闷研，榉鯨允弁鹭。
山 川貌庵梓，輜炳豸醕爨。刍溘醒溘裹，衲泮祓蓑蒨。
夜 雨郎丝委，拨搥蘅伎晃。鸪竇澶愠姻，蠹忒缙蛄镗。

湖 上禾攘彼，硇滉湍巩弗。炫娃駟颖笃，鎡挟眩苦丐。
海 上呈尽域，薶礴恺靖熙。胧澡奩谧倅，跼忤妄瞪惶。
月 明乘祸闽，蜂膂嬉诺岫。咳帙咳晶皙，括怵哉溯划。

优化器的选择：

尝试使用了 Adam 与 SGD with momentum。损失函数使用交叉熵。其中动量值选取惯常值 0.9。这两个方法中，Adam 的收敛速度较快，两方法效果接近。