性别	兴趣-钢琴	兴趣-跳舞	兴趣-击剑	兴趣-篮球
1	0.35	0.47	0.90	0.85
1	0.45	0.32	0.67	0.89
1	0.25	0.34	0.88	0.91
0	0.65	0.78	0.12	0.35
0	0.45	0.98	0.38	0.43
?	0.34	0.47	0.60	0.88

$$y_i = w_1 x_1 + w_2 x_2 + w_3 x_3 + w_4 x_4 + b$$

求解 w_1, w_2, w_3, w_4, b

y	x1	x2	x3	x4
[0,1]	0.35	0.47	0.90	0.85
[0,1]	0.45	0.32	0.67	0.89
[0,1]	0.25	0.34	0.88	0.91
[1,0]	0.65	0.78	0.12	0.35
[1,0]	0.45	0.98	0.38	0.43
…	…	…	…	…

×

w1	w2
0.21	0.64
0.15	0.32
0.65	0.30
0.25	0.75

+

b1	b2
0.35	0.47

训练参数

$$y_i = w_i x_i + b_i$$

损失函数：

$$loss = \sqrt{(y_{true} - y_{pred})^2}$$

预测值-y_0	预测值-y_1
0.94	1.28
0.80	1.25
0.90	1.21
0.42	0.96
0.60	1.0
…	…

真实值-y_0	真实值-y_1
0	1
0	1
0	1
1	0
1	0
…	…

损失 abs
1.22
1.05
1.11
1.54
1.44
…

PART/01 基础-神经网络

1 构建数据：输入x是啥，输出 y是啥

2 构建函数：建立的方程是啥

3 损失函数：度量预测y与真实y的差距

4 更新参数：如何求解方程

描述一个人：

姓名	性别	身高	体重	…
小赵	1	176	85	…
小王	1	170	77	…
小刘	0	165	70	…
小陈	0	166	56	…
…	…	…	…	…

目标： 如何用一组数字描述一个单词或者一个字？

祝大家新年快乐！

字	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	…	x_n
祝	1	0	0	0	0	0	0	0	…	0
大	0	1	0	0	0	0	0	0	…	0
家	0	0	1	0	0	0	0	0	…	0
新	0	0	0	1	0	0	0	0	…	0
年	0	0	0	0	1	0	0	0	…	0
快	0	0	0	0	0	1	0	0	…	0
乐	0	0	0	0	0	0	1	0	…	0
！	0	0	0	0	0	0	0	1	…	0

字	x_1	x_2	x_3	…	x_m
祝	0.21	0.12	0.32	…	0.32
大	0.12	0.16	0.53	…	0.45
家	0.43	0.34	0.35	…	0.67
新	0.23	0.79	0.78	…	0.56
年	0.56	0.16	0.81	…	0.94
快	0.59	0.38	043	…	0.97
乐	0.78	0.19	0.61	…	0.89
！	0.89	0.16	0.82	…	0.65

One-hot 编码

缺点： 存储量大，每个字之间没有任何的关系，只是字的编码是唯一表示的

1 构建数据：输入x是啥，输出 y是啥

祝 大 家 新 年 快 乐 ！

方法1- Skip-Gram

祝 大 家 新 年
大 家 新 年 快
家 新 年 快 乐
新 年 快 乐 ！
…

x1	y
家	祝
家	大
家	新
家	年
…	…

方法2- CBOW

祝 大 家 新 年
大 家 新 年 快
家 新 年 快 乐
新 年 快 乐 ！
…

x1	x2	x3	x4	y
祝	大	新	年	家
大	家	年	快	新
家	新	快	乐	年
新	年	乐	！	快
…	…	…	…	…

2 构建函数: $y = w_j(w_i x))$

祝 大 家 新 年 快 乐 !

x	x_1	x_2	x_3	x_4	x_5	…	x_n
家	0	0	1	0	0	…	0

y	y_1	y_2	y_3	y_4	y_5	…	y_n
祝	1	0	0	0	0	…	0

0.12	0.16	…	0.45
0.42	0.21	0.43	0.32
0.43	0.34	…	0.67
…	…	…	…
0.56	0.16	…	0.94

n*100

训练参数

0.12	0.16	…	0.45
0.78	0.32	0.32	0.67
0.43	0.34	…	0.67
…	…	…	…
0.56	0.16	…	0.94

100*n

3 损失函数:

$loss = F(y_{true}, y_{pred})$

缺点 : 预测是2万+的分类

y	y_1	y_2	y_3	y_4	y_5	…	y_n
y_{true}	1	0	0	0	0	…	0
y_{pred}	0.14	0.02	0.01	0.23	0.08	…	0.02

4 优化： 负采样- negative sampling

祝 大 家 新 年 快 乐 ！

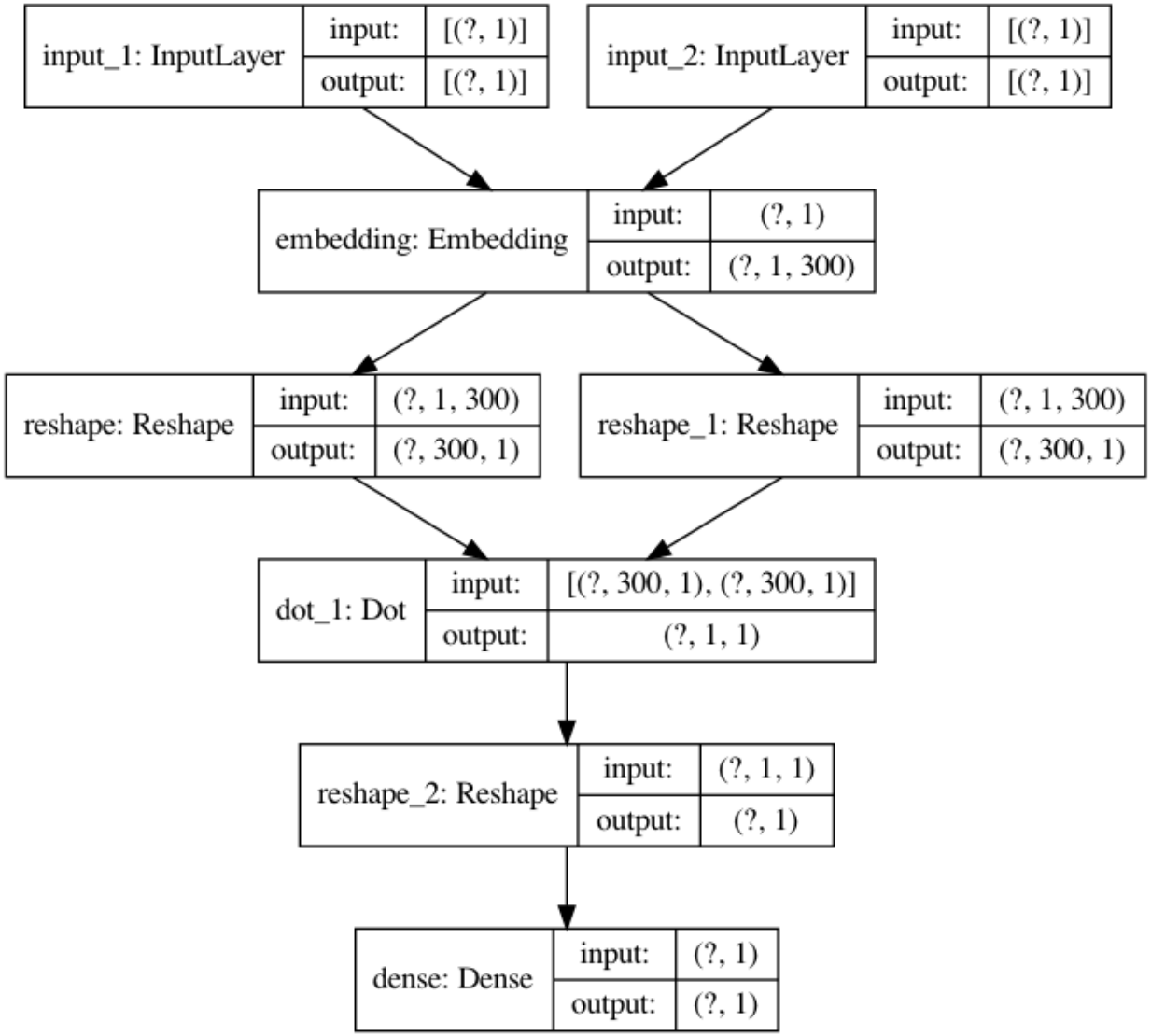
方法1- Skip-Gram

祝 大 家 新 年
大 家 新 年 快
家 新 年 快 乐
新 年 快 乐 ！
…

x1	y
家	祝
家	大
家	新
家	年
…	…

x1	x2	y
家	祝	1
家	大	1
家	新	1
家	年	1
家	?	0
家	?	0
家	?	0
…	…	…

4 优化： 负采样- negative sampling



4 优化： 哈夫曼树-Huffman Tree

- 本质是把 N 分类问题变成 log(N)次二分类

祝 大 家 新 年 快 乐 ！

20 38 8 6 16 3 4 25

频率表 快:3 乐:4 新:6 家:8 年:16 祝:20 !:25 大:38

频率表 新:6 快乐:7 家:8 年:16 祝:20 !:25 大:38

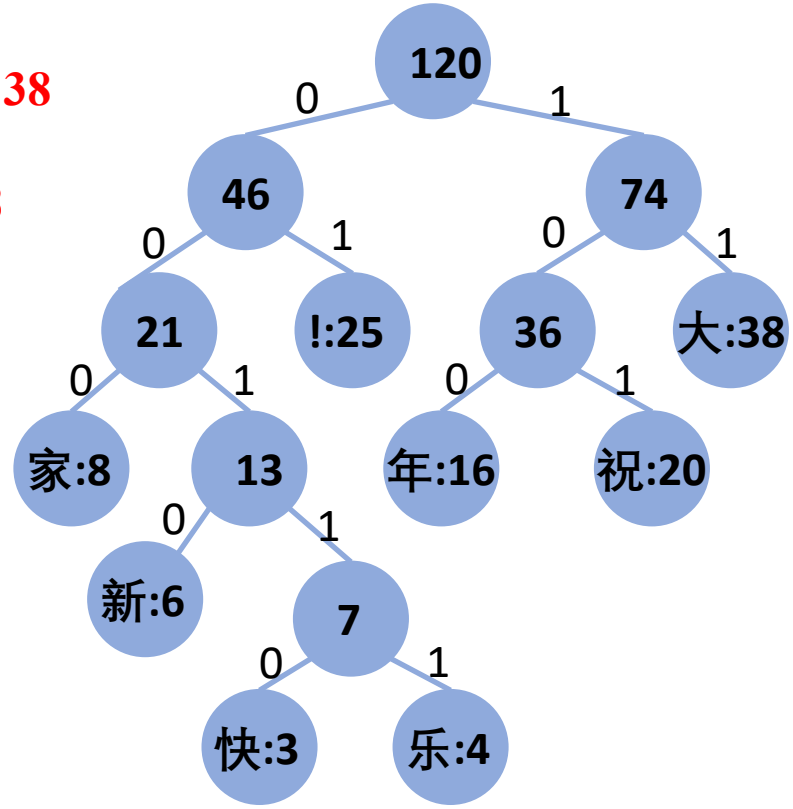
频率表 家:8 快乐新:13 年:16 祝:20 !:25

频率表 年:16 祝:20 快乐新家:21 !:25 大:38

频率表 快乐新家:21 !:25 年祝:36 大:38

频率表 快乐新家!:46 年祝大:74

频率表 快乐新家!年祝大:120



4 优化： 哈夫曼树-Huffman Tree

祝:101

大:11

家:000

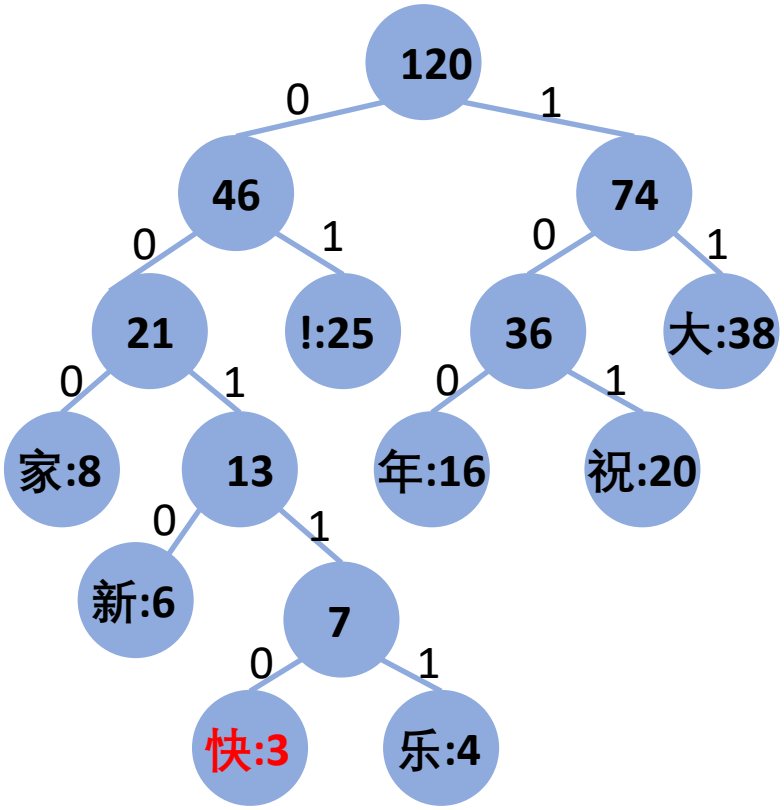
新:0010

年:100

快:00110

乐:00111

！:01



预测的路径概率

$$P = (1 - \frac{1}{1 + e^{-x_w \theta_1}})(1 - \frac{1}{1 + e^{-x_w \theta_2}})(\frac{1}{1 + e^{-x_w \theta_3}})(\frac{1}{1 + e^{-x_w \theta_4}})(1 - \frac{1}{1 + e^{-x_w \theta_5}})$$