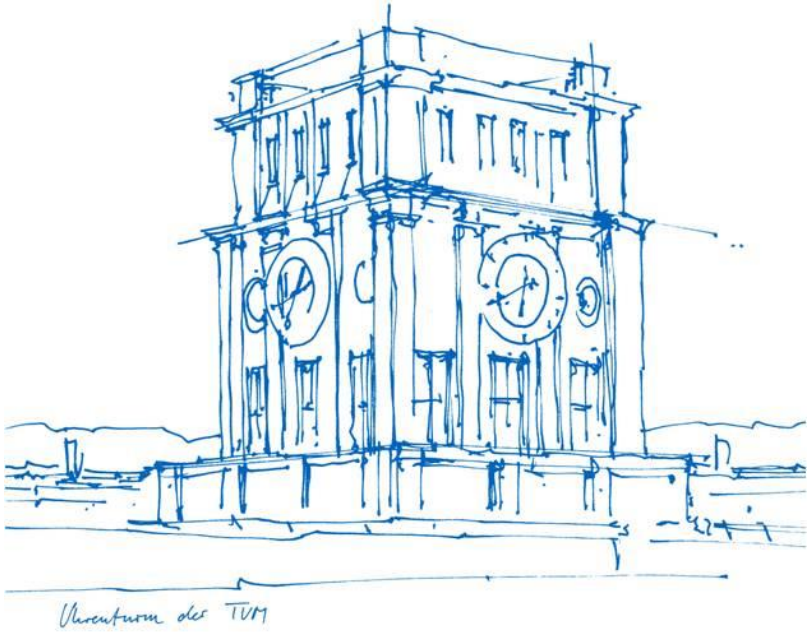
Multi-Lingual Theme Prediction of Customer Reviews Using Deep Pre-Trained Embeddings

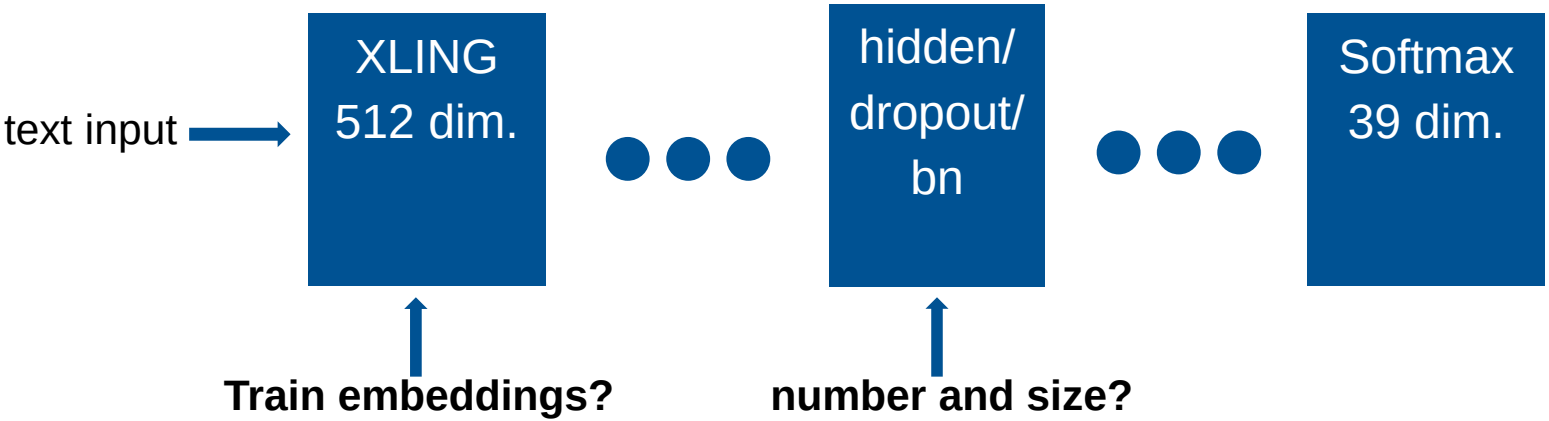
Team 06

Michael Sorg

05.06.2019

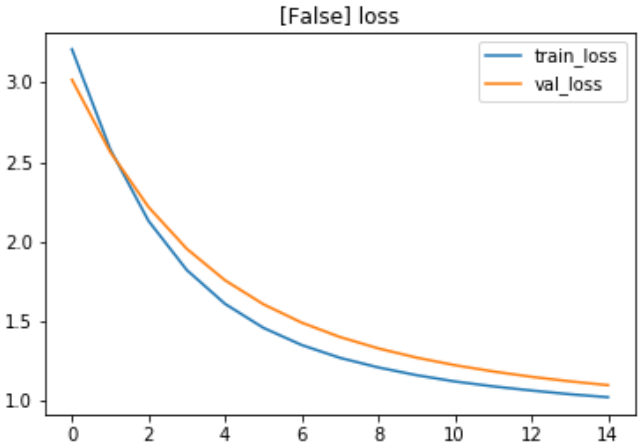


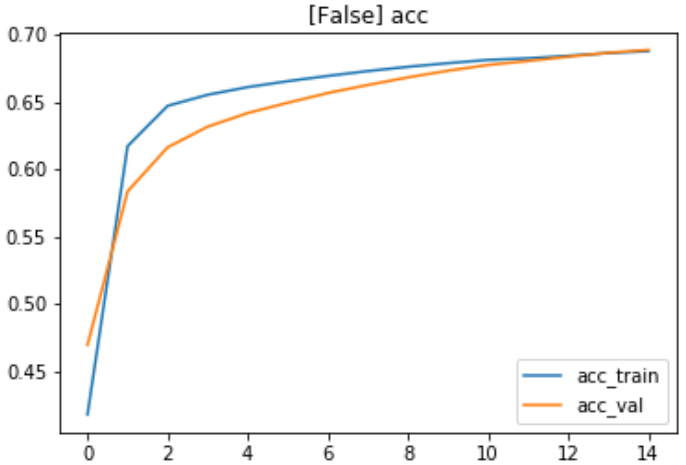
Network search



Baseline experiments

- Amazon Multilingual UK with 100k examples
- One hidden layer (only xling + softmax)

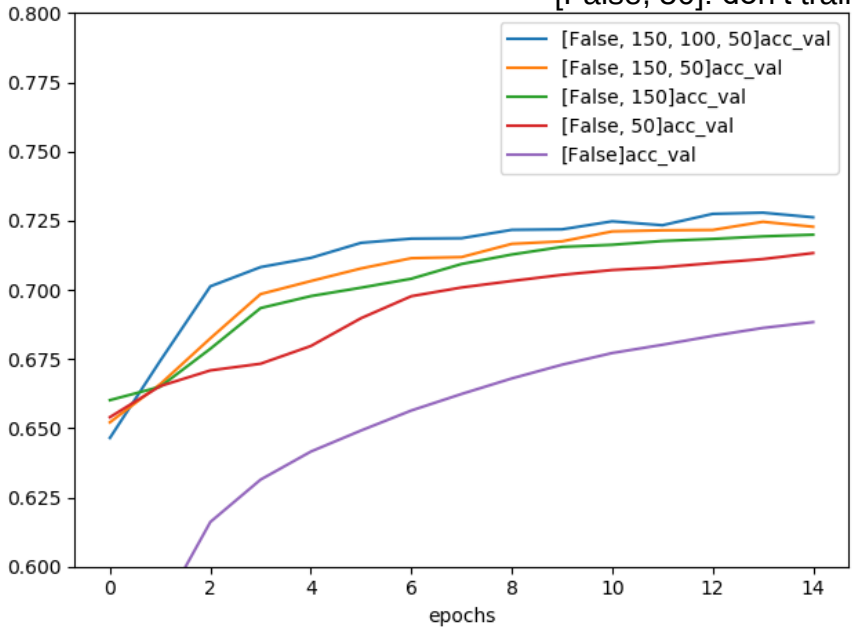




Architecture search

- Amazon Multilingual UK with 100k examples

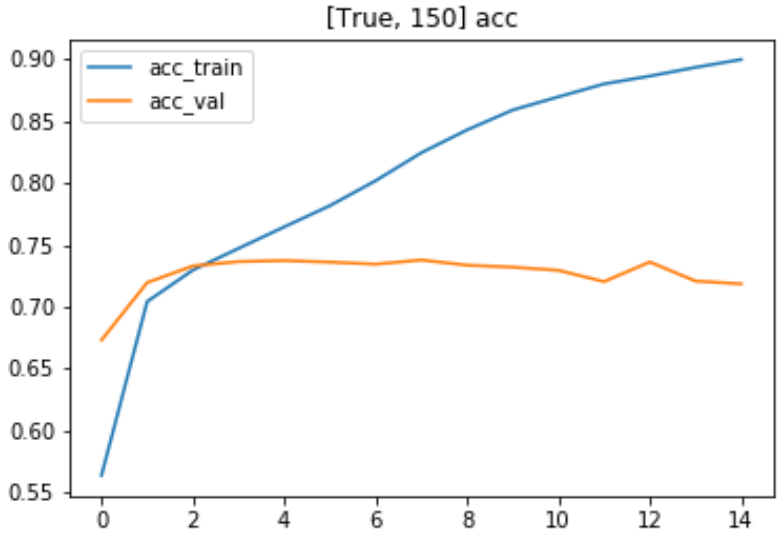
[False, 50]: don‘t train embeddings, fc with 50 units + softmax



Architecture search

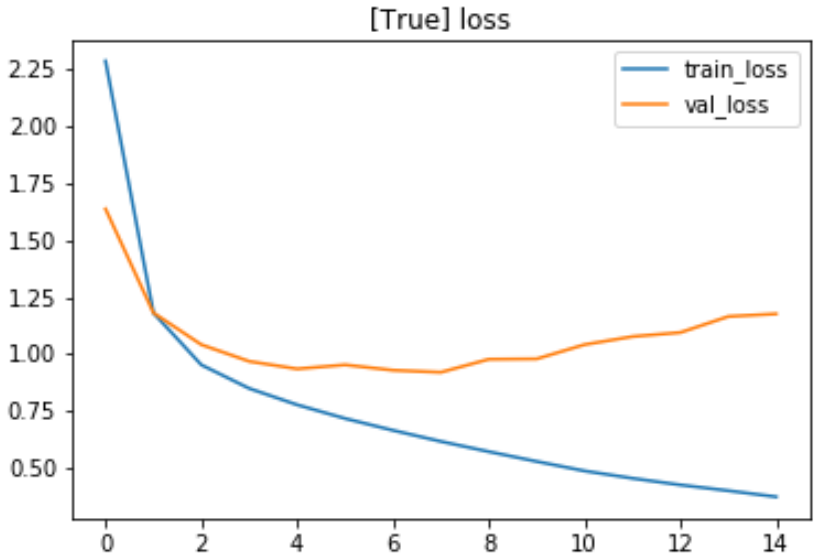
```
self.train_op = tf.train.AdamOptimizer(epsilon=0.1).minimize(self.loss)
```

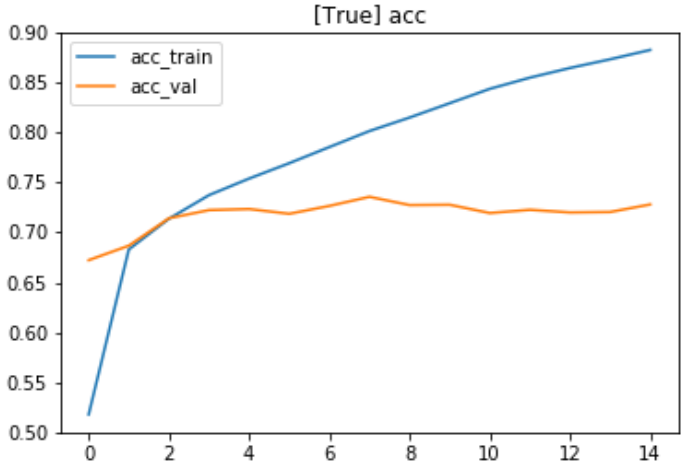
- Train also xling embeddings (100k samples)



Architecture search

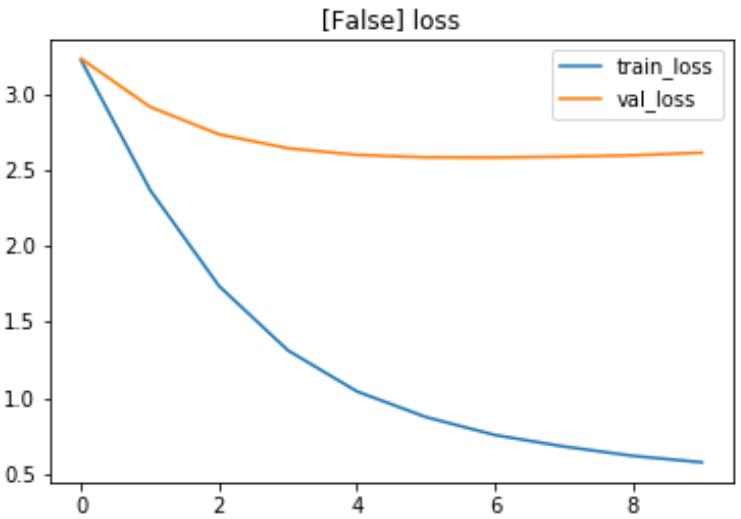
- Train also xling embeddings
- Add dropout to prevent overfitting

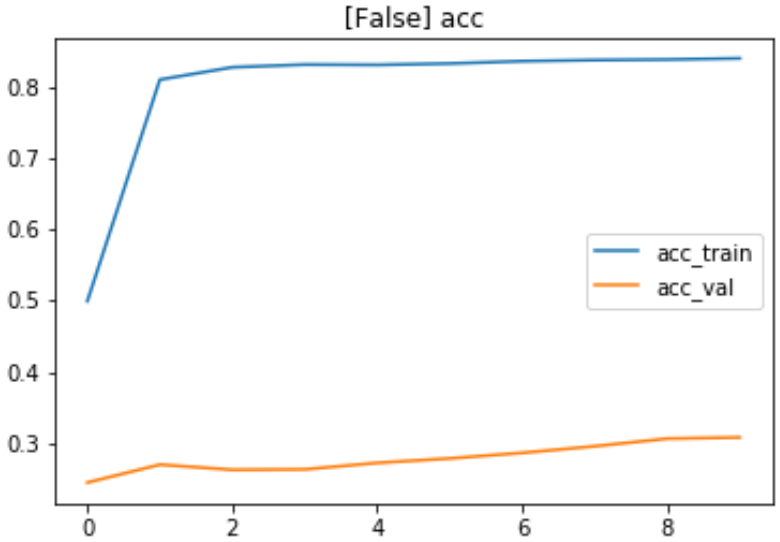




Architecture search

- Amazon Multilingual **US** with 100k examples





```
de.product_category.value_counts()
```

```
Video DVD                  279068
Music                      160588
Books                       63784
Mobile_Apps                 54709
Digital_Video_Download      25124
Digital_Music_Purchase      21554
Toys                        18602
Digital_Ebook_Purchase      12872
PC                          12250
Camera                       5421
Wireless                     4441
Electronics                  4035
Video                        2927
Sports                       2034
Video Games                  1706
Watches                      1575
Shoes                        1517
Home                         1454
Musical Instruments          1094
Baby                          810
Home Improvement              672
Home Entertainment            605
Office Products               412
Personal_Care_Appliances      411
Automotive                    410
Lawn and Garden               397
Luggage                       247
Kitchen                       120
Furniture                      93
Health & Personal Care         37
Software                       19
Pet Products                    2
Grocery                         2
Beauty                          1
Name: product_category, dtype: int64
```

```
[9] us.product_category.value_counts()
```

```
Mobile_Apps                1467128
Digital_Ebook_Purchase     1242173
Video DVD                  1093612
Digital_Video_Download     1051622
Books                       836136
Music                       776810
Digital_Music_Purchase      107461
Toys                         57465
PC                           56817
Video                        46697
Home Entertainment           36298
Wireless                     22637
Camera                       16911
Video Games                  15398
Electronics                  11421
Musical Instruments          10914
Watches                      10537
Tools                         7475
Shoes                         7342
Baby                          5868
Sports                        4123
Home Improvement              3698
Outdoors                      3184
Office Products               2306
Home                          1998
Kitchen                       1834
Lawn and Garden               1204
Health & Personal Care        1066
Automotive                     223
Mobile_Electronics             184
Apparel                        121
Luggage                         78
Beauty                          52
Software                        52
Grocery                         18
Personal_Care_Appliances         9
Furniture                        8
Pet Products                     5
2012-12-22                       1
Name: product_category, dtype: int64
```

```
f = us[0:100000]
f.product_category.value_counts()
```

```
Music              43422
Books              23948
Video              17702
Video DVD          14911
Toys                  10
Office Products        5
Tools                  2
Name: product_category, dtype: int64
```

```
f2 = de[0:100000]
f2.product_category.value_counts()
```

```
Video DVD                  40660
Music                      17961
Mobile_Apps                11905
Digital_Video_Download      6387
Books                       5727
Toys                        3633
Digital_Music_Purchase      3609
Digital_Ebook_Purchase      2579
PC                          2396
Wireless                     876
Camera                       810
Electronics                  779
Video Games                  404
Watches                      324
Shoes                        301
Sports                       267
Musical Instruments          263
Home                         244
Baby                         177
Home Entertainment           133
Home Improvement             124
Automotive                    80
Office Products               78
Lawn and Garden               61
Personal_Care_Appliances      61
Video                         55
Luggage                       51
Furniture                     28
Kitchen                       16
Health & Personal Care         9
Software                       2
Name: product_category, dtype: int64
```

How to find architecture?

- Use more advanced search techniques (Progressive Neural Architecture Search (PNAS), Efficient Neural Architecture Search (ENAS), Reinforcement learning)

- Tradeoff between time and size of training set ?

Infrastructure Issues

- Training for more than 30 min. fails in most cases