

Happiness narratives

- Data Discrepancies
 - The stats
- Results
 - SVR Joy
 - SVM Joy
 - Lasso Joy
 - Lasso Sadness
 - SVR Sadness
 - SVM Sadness
- Correlation
- Neural Networks
 - RELU unscaled results
 - Linear Unscaled results
 - RNN scaled results
 - RNN unscaled results
- Meeting Notes
 - 5-13

Data Discrepancies

There's more data used for the hedometer evaluation than there is for the ml training and evaluation.

The stats

The Hedometer data has 3656 entries The ML data has 3343 enttries.

The preprocessing done on the hedometer data makes it difficult to compare the two files.

Lowercasing, all punctuation removed, spaced injeted.

Results

SVR Joy

ROC - AUC : 0.5053219775610537

SVM Joy

```
MSE: 0.6431352807865082
Mean Absolute Error 0.6158711064947314
R2: 0.2920088534366445
Pearson correlation: (0.5917811508024919, 0.0)
0 values were squeezed into the score range0 values were squeezed into the score
range
      precision    recall  f1-score   support
1  0.38461538  0.06944444  0.11764706      288
0.47570621      883
3  0.49695122  0.62837317  0.55498808     1297
0.53450808     1010
5  0.37142857  0.07692308  0.12745098     169
accuracy              0.49766932     3647  macro avg  0.45123727  0.35842369
0.36206008     3647
weighted avg  0.48561312  0.49766932  0.47577265     3647
[[ 20 172  80  16   0]
 [ 25 421 346  91   0]
 [  6 229 815 246   1]
 [  1  62 380 546  21]
 [  0   3  19 134  13]]
ROC - AUC : 0.6071926695107926
```

Lasso Joy

ROC - AUC : 0.5058964520950757

Lasso Sadness

.....

SVR Sadness

```
SE: 0.9454209231533085
Mean Absolute Error 0.8007222131105772
R2: -0.05726754954823532
Pearson correlation: (0.191833119260266, 3.3651319183614605e-32)1 values were squeezed into
the score range0 values were squeezed into the score range
/home/kenneth/venvs/crfsuite/lib/python3.8/site-
packages/sklearn/metrics/_classification.py:1272: UndefinedMetricWarning: Precision and F-
score are ill-defined and being set to 0.0 in labels with no predicted samples. Use
`zero_division` parameter to control this behavior.  _warn_prf(average, modifier, msg_start,
len(result))

      precision    recall  f1-score   support

1  0.00000000  0.00000000  0.00000000        146          2  0.06111111  0.01674277
0.02628435          657

3  0.32739130  0.60725806  0.42542373        1240          4  0.41384996  0.38704703
0.40000000          1297

5  0.00000000  0.00000000  0.00000000          384

accuracy              0.33995704        3724  macro avg  0.16047047  0.20220957
0.17034162          3724

weighted avg  0.26393088  0.33995704  0.28560533        3724

[[ 0  4 111  31  0]
 [ 0 11 497 148  1]
 [ 0 136 753 342  9]
 [ 0  27 747 502 21]
 [ 0  2 192 190  0]]

ROC - AUC : 0.5030923629109962
```

SVM Sadness

ROC - AUC : 0.578488077944395

Correlation

Binyan was doing pearson's correlation. Thinks maybe we should do spearman

Neural Networks

RELU unscaled results

Joy train on 9 folds, test on 1

```
2020-05-13 12:15:50,734 - INFO - allennlp.common.util - Metrics: {
  "best_epoch": 12,
  "peak_cpu_memory_MB": 2626.072,
  "peak_gpu_0_memory_MB": 8085,
  "peak_gpu_1_memory_MB": 21478,
  "training_duration": "0:23:41.867688",
  "training_start_epoch": 0,
  "training_epochs": 21,
  "epoch": 21,
  "training_pearson": 0.9939782728494088,
  "training_mae": 0.14345509548909208,
  "training_loss": 0.03389307007519076,
  "training_cpu_memory_MB": 2626.072,
  "training_gpu_0_memory_MB": 7409,
  "training_gpu_1_memory_MB": 18680,
  "validation_pearson": 0.8524911266953036,
  "validation_mae": 0.7244842611554498,
  "validation_loss": 0.8840915312369665,
  "best_validation_pearson": 0.8559774687113151,
  "best_validation_mae": 0.6610433984638224,
  "best_validation_loss": 0.7440112556020418
}
```

Sadness train on 9 folds,

test on 1

These are longer and they're causing clipping. I don't know how many are causing clipping though because allennlp only reports the first case of clipping.

```
2020-05-13 12:22:49,449 - INFO - allennlp.common.util - Metrics: {
  "best_epoch": 1,
  "peak_cpu_memory_MB": 2759.188,
  "peak_gpu_0_memory_MB": 7409,
  "peak_gpu_1_memory_MB": 18680,
  "training_duration": "0:17:02.438299",
  "training_start_epoch": 0,
  "training_epochs": 10,
  "epoch": 10,
  "training_pearson": -0.01633020200422353,
  "training_mae": 1.3776687749682646,
  "training_loss": 2.781699788029837,
  "training_cpu_memory_MB": 2759.188,
  "training_gpu_0_memory_MB": 7409,
  "training_gpu_1_memory_MB": 11,
  "validation_pearson": 0,
  "validation_mae": 1.355329878786777,
  "validation_loss": 2.7624370823515223,
  "best_validation_pearson": 0.23847302176509885,  "best_validation_mae":
1.3548242284896526,
  "best_validation_loss": 2.7604094781774156
}
```

Linear Unscaled results

Joy

$lr = 0.001$

Really bad, mean absolute error of 1.3ish at the end.

$lr = 0.0001$

Seemed to overfit, learning rate continued to decrease. for training but not validation.

```
2020-05-14 18:53:12,495 - INFO - allennlp.common.util - Metrics: {
  "best_epoch": 11,
  "peak_cpu_memory_MB": 2630.66,
  "peak_gpu_0_memory_MB": 1,
  "peak_gpu_1_memory_MB": 21478,
  "training_duration": "0:23:01.688213",
  "training_start_epoch": 0,
  "training_epochs": 20,
  "epoch": 20,
  "training_pearson": 0.9952644629823881,
  "training_mae": 0.12786512176923553,
  "training_loss": 0.02625432804009868,
  "training_cpu_memory_MB": 2630.66,
  "training_gpu_0_memory_MB": 1,
  "training_gpu_1_memory_MB": 18680,
  "validation_pearson": 0.8548239304162555,
```

```
"validation_mae": 0.6628524492371757,  
"validation_loss": 0.7741623421510061,  
"best_validation_pearson": 0.8563759958887212,  
"best_validation_mae": 0.6515413680166569,  
"best_validation_loss": 0.7303319076697031  
}
```

lr = 0.00001

```
2020-05-14 21:39:30,893 - INFO - allennlp.common.util - Metrics: {  
  "best_epoch": 2,  
  "peak_cpu_memory_MB": 2627.62,  
  "peak_gpu_0_memory_MB": 1,  
  "peak_gpu_1_memory_MB": 21478,  
  "training_duration": "0:13:09.258866",  
  "training_start_epoch": 0,  
  "training_epochs": 11,  
  "epoch": 11,  
  "training_pearson": 0.9709087494155519,  
  "training_mae": 0.31317798209277703,  
  "training_loss": 0.15988704243909965,  
  "training_cpu_memory_MB": 2627.62,  
  "training_gpu_0_memory_MB": 1,  
  "training_gpu_1_memory_MB": 18094,  
  "validation_pearson": 0.8498408919748213,  
  "validation_mae": 0.721538697934215,  
  "validation_loss": 0.8732728213071823,  
  "best_validation_pearson": 0.8493931180602585,  
  "best_validation_mae": 0.6772722928029187,  
  "best_validation_loss": 0.7859473476807276  
}
```


RNN scaled results

3 hid

lr 0.00001

```
2020-05-14 22:57:02,171 - INFO - allennlp.common.util - Metrics: {  
  "best_epoch": 28,  
  "peak_cpu_memory_MB": 1997.92,  
  "peak_gpu_0_memory_MB": 2697,  
  "peak_gpu_1_memory_MB": 11,  
  "training_duration": "0:06:15.555519",  
  "training_start_epoch": 0,  
  "training_epochs": 37,  
  "epoch": 37,  
  "training_pearson": 0.8835897766720373,  
  "training_mae": 0.6050379267542354,  
  "training_loss": 0.012446582688072931,  
  "training_cpu_memory_MB": 1997.892,  
  "training_gpu_0_memory_MB": 2697,  
  "training_gpu_1_memory_MB": 11,  
  "validation_pearson": 0.7651979545423095,  
  "validation_mae": 0.854339599609375,  
  "validation_loss": 0.02388349245302379,  
  "best_validation_pearson": 0.763310168046131,  
  "best_validation_mae": 0.8508703021026365,  
  "best_validation_loss": 0.023334396770223975  
}
```

lr 0.0001

```
2020-05-14 23:00:13,281 - INFO - allennlp.common.util - Metrics: {
  "best_epoch": 3,
  "peak_cpu_memory_MB": 1984.856,
  "peak_gpu_0_memory_MB": 2437,
  "peak_gpu_1_memory_MB": 11,
  "training_duration": "0:02:09.181165",
  "training_start_epoch": 0,
  "training_epochs": 12,
  "epoch": 12,
  "training_pearson": 0.9309341933147448,
  "training_mae": 0.4662433107257326,
  "training_loss": 0.007578380315483195,
  "training_cpu_memory_MB": 1984.816,
  "training_gpu_0_memory_MB": 2437,
  "training_gpu_1_memory_MB": 11,
  "validation_pearson": 0.7313824688523022,
  "validation_mae": 0.8880522283261034,
  "validation_loss": 0.028224910454203684,
  "best_validation_pearson": 0.7701748503103649,
  "best_validation_mae": 0.8408898201914168,
  "best_validation_loss": 0.022758180275559425
}
```

2 hid

RNN unscaled results

Glove embeddings

64 dimensional RNN

1 hidden layer

2 hidden layers

0.0001 lr

```
2020-05-19 23:19:25,110 - INFO - allennlp.common.util - Metrics: {  
  "best_epoch": 35,  
  "peak_cpu_memory_MB": 1715.32,  
  "peak_gpu_0_memory_MB": 1,  
  "peak_gpu_1_memory_MB": 636,  
  "training_duration": "0:01:44.572271",  
  "training_start_epoch": 0,  
  "training_epochs": 44,  
  "epoch": 44,  
  "training_pearson": 0.8755669240062097,  
  "training_mae": 0.6300580370557177,  
  "training_loss": 0.6342626507793154,
```

```
"training_cpu_memory_MB": 1715.3,  
"training_gpu_0_memory_MB": 1,  
"training_gpu_1_memory_MB": 636,  
"validation_pearson": 0.7133927439166854,  
"validation_mae": 0.913505204604321,  
"validation_loss": 1.405275821685791,  
"best_validation_pearson": 0.7112064102524507,  
"best_validation_mae": 0.9231068901617251,  
"best_validation_loss": 1.3673358609278996  
}
```

3 hidden layers

Meeting Notes

Meeting Notes

5-13

Variety of different conferences and journals we could submit to.

Digital Scholarship in the Humanities is a DH journal we could submit to.