Coursework Assignment B - 2020 Code ▾

CS4125 Seminar Research Methodology for Data Science

By- Aditya Kunar (5074274), Sharwin Bobde (5011639), Pavel Hoogland (4450892)

Research Question 1

Question - How much does transfer learning improve over typical non-transfer learning?

1) Understanding the dataset.

a) Counts for the observations for each model.

```
 B1  B2  B3  M1  M2  M3  MF  MN   S
  7   7   7 546 539 504 560 490 336
```

Here we see that we only have 21 points for the first scenario(B1,B2,B3) which involves no transfer learning. Whereas with transfer learning, there are many more data points.

b) Observing the mean scores for the models.

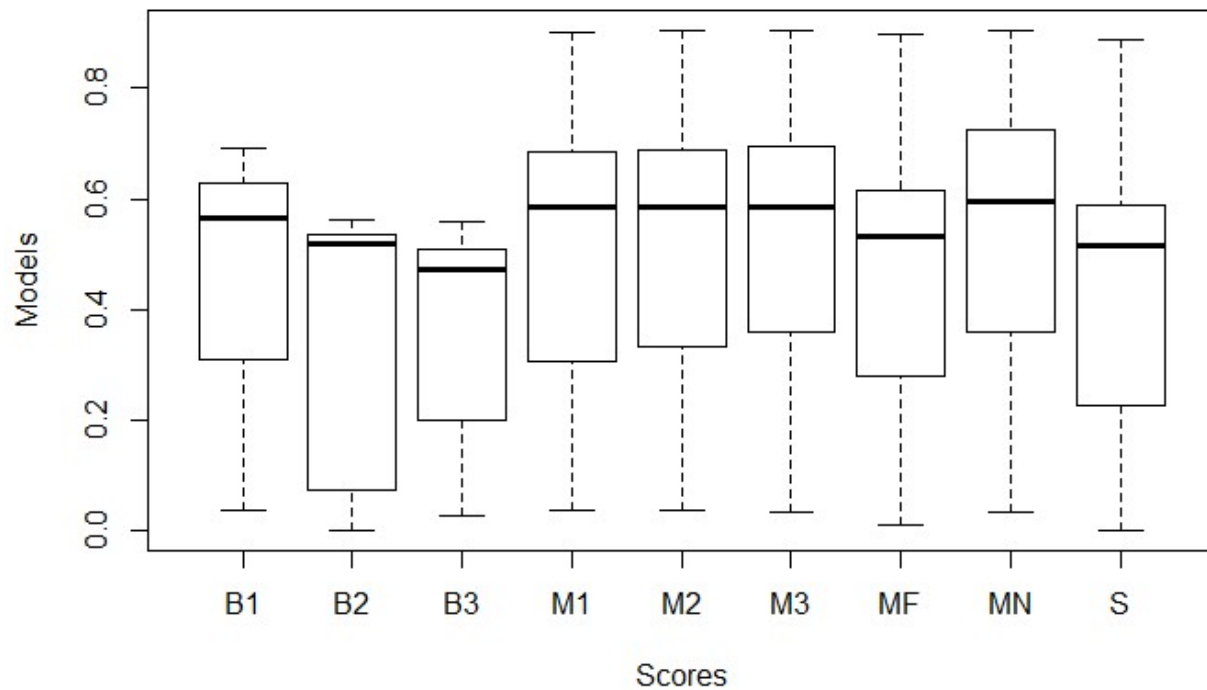| model	score
<fctr>	<dbl>
B1	0.4528788
B2	0.3278496
B3	0.3534039
M1	0.5088922
M2	0.5146289
M3	0.5204925
MF	0.4546982
MN	0.5310110
S	0.4287001

> 9 rows

We can see that the MN model has the highest mean score(.53) whereas the B2 model has the lowest mean score(0.32). We can also start to see that transfer learning at least doesn't negatively impact the mean scores of the models. However we still need to determine if anything can be said about it postively impacting the scores of the models in our dataset.

c) Observing the means for the models.

TeD <fctr>	score <dbl>
TeD1	0.78805072
TeD2	0.57703670
TeD3	0.69891817
TeD4	0.48141839
TeD5	0.05004097
TeD6	0.60545365
TeD7	0.26725158
7 rows	

Here we see the scores aggregated on the test datasets. We know that TeD1 to TeD4 are classification datasets and their mean scores are found to be higher than for the other tasks such as regression and recommendation. Infact we see the lowest performance for the recommendation task dataset i.e TeD5 with a mean score of 0.05. We are beginning to realise that the *mean score of any model(transfer learning or not) might also be impacted by the test dataset as well*.

d) Visualisations of model and dataset scores.

By looking at the box plots of scores aggregated on models, we can see that models M1, M2, M3 and MN have somewhat similar scores and the median scores of these models tend to be higher than for models B1,B2 and B3 for which transfer learning has not been used. We can also clearly see that the models for which transfer learning has been used have a larger spread in their scores. From this boxplot we can start to appreciate that models play a role in affecting the scores in our dataset.
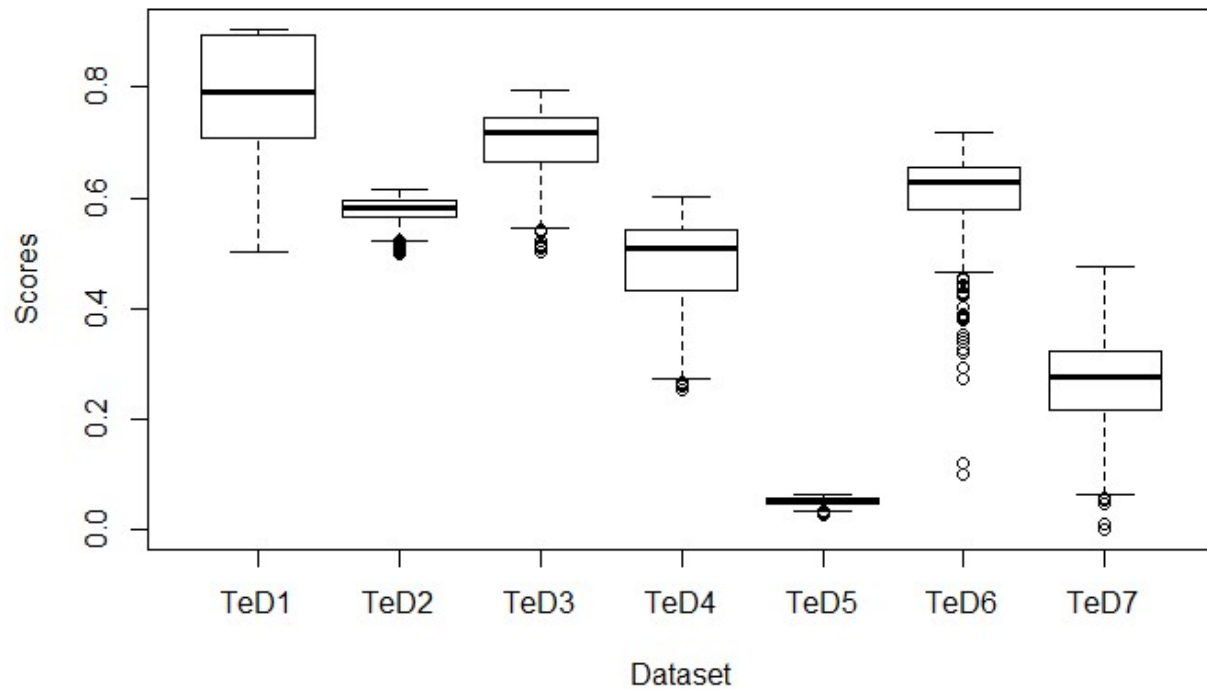
By looking at scores aggregated on test datasets, we can clearly see there is no homogeneity amongst the scores on different test datasets. We see that TeD1 has the highest median score and TeD5 has the lowest with a very small interquartile range as well. We also notice that TeD6 has quite a few outliers as compared to the other test datasets. But it is clear that test datasets do have an influence on the scores in our dataset.

e) Visualising the interaction between model and test datasets.
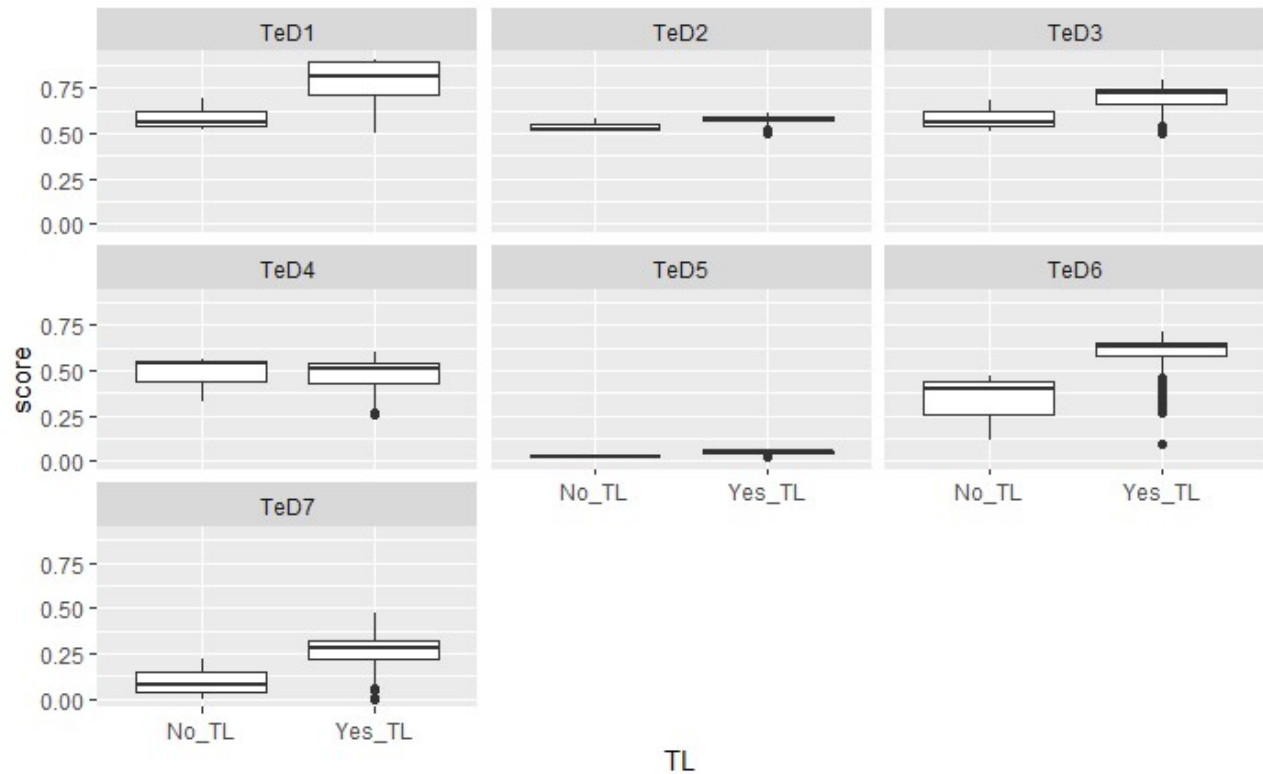
From the above visualization, we can see that the test dataset creates a great degree of variation in the differences between models scores with transfer learning and without. We can also notice that except for TeD4 and TeD5, models with transfer learning tend to do better on most other test datasets.

2) Mutli Level Modelling

We wanted to perform a hierarchical analysis due to the fact that there was a clear dependency in observations obtained from the same test dataset. Therefore, we have used a multi-level model wherein the random effects comprise of random interecepts for the different test datasets and random slopes for the difference in scenarios between whether transfer learning was used or not used. The fixed effect of our model solely consists of a predictor variable indicating whether transfer learning was used or not

```
Linear mixed-effects model fit by maximum likelihood
  Data: data
```

AIC <dbl>	BIC <dbl>	logLik <dbl>
-7251.012	-7214.981	3631.506

1 row

```
Random effects:
 Formula: ~TL | TeD
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev      Corr
(Intercept) 0.21160585 (Intr)
TLYes_TL    0.08640309 0.123
Residual    0.07116947

Fixed effects: score ~ TL
```

	Value	Std.Error	DF	t-value	p-value
	<chr>	<chr>	<chr>	<chr>	<chr>
(Intercept)	0.3780441	0.08150061	2988	4.638543	0.0000
TLYes_TL	0.1182376	0.03619767	2988	3.266441	0.0011

2 rows

```
 Correlation:
        (Intr)
TLYes_TL 0.027

Standardized Within-Group Residuals:
       Min          Q1         Med          Q3         Max
-7.14526038 -0.42188499  0.07890215  0.61218608  2.90748161

Number of Observations: 2996
Number of Groups: 7
```

Based on this model, we see that the presence of information regarding the use of transfer learning(t=3.26,df=2988,p.<0.01) has a significant impact on the scores of our models. We see that there is on average, a positive deviation of 0.12(95% CI, 0.04 to 0.18) across the different datasets for when transfer learning is used as compared to when it's not used.

```
Approximate 95% confidence intervals

 Fixed effects:
                 lower       est.       upper
(Intercept) 0.21829445 0.3780441 0.5377937
TLYes_TL    0.04728638 0.1182376 0.1891887
attr(,"label")
[1] "Fixed effects:"

 Random Effects:
  Level: TeD
                              lower        est.       upper
sd((Intercept))          0.12268074 0.21160585 0.3649883
sd(TLYes_TL)             0.04543044 0.08640309 0.1643280
cor((Intercept),TLYes_TL) -0.74950543 0.12337320 0.8396056

 Within-group standard error:
     lower       est.       upper
0.06938594 0.07116947 0.07299883
```

From the results shown above it can be seen that there is a significant variation in the scores caused by differences in the test datasets. This can be verified by looking at the confidence intervals of the random effects and more specifically the standard deviation between the fitted intercepts for the different test datasets which is 0.21 (95% CI, 0.12 to 0.36). Futhermore, we can also see that the standard deviation between the slopes is 0.086 (95% CI, 0.04 to 0.16) is not too big which is fitted on the basis of whether transfer learning was used for different test datasets suggesting that the different test datasets mostly had a beneficial effect from the use of transfer learning. Lastly, we can see that the confidence intervals are quite broad, this is an indication that there is a great deal of uncertainty in our estimates and we believe this is due to the imbalance of data points for when transfer learning was used vs when it wasn't as shown in the very beginning of the analysis.

The use of EMMEANS.

For more clarity in our interpretation, we can also use the library emmeans.

```
$emmeans
TeD = TeD1:
 TL      emmean      SE    df lower.CL upper.CL
 No_TL  0.5900 0.04109 2982   0.5095   0.6706
 Yes_TL 0.7894 0.00345 2982   0.7827   0.7962

TeD = TeD2:
 TL      emmean      SE    df lower.CL upper.CL
 No_TL  0.5379 0.04109 2982   0.4574   0.6185
 Yes_TL 0.5773 0.00345 2982   0.5705   0.5841

TeD = TeD3:
 TL      emmean      SE    df lower.CL upper.CL
 No_TL  0.5829 0.04109 2982   0.5023   0.6634
 Yes_TL 0.6997 0.00345 2982   0.6930   0.7065

TeD = TeD4:
 TL      emmean      SE    df lower.CL upper.CL
 No_TL  0.4790 0.04109 2982   0.3985   0.5596
 Yes_TL 0.4814 0.00345 2982   0.4747   0.4882

TeD = TeD5:
 TL      emmean      SE    df lower.CL upper.CL
 No_TL  0.0302 0.04109 2982  -0.0503   0.1108
 Yes_TL 0.0502 0.00345 2982   0.0434   0.0569

TeD = TeD6:
 TL      emmean      SE    df lower.CL upper.CL
 No_TL  0.3299 0.04109 2982   0.2493   0.4104
 Yes_TL 0.6074 0.00345 2982   0.6006   0.6142

TeD = TeD7:
 TL      emmean      SE    df lower.CL upper.CL
 No_TL  0.0964 0.04109 2982   0.0158   0.1769
 Yes_TL 0.2685 0.00345 2982   0.2617   0.2752

Confidence level used: 0.95

$contrasts
TeD = TeD1:
 contrast        estimate     SE    df t.ratio p.value
 No_TL - Yes_TL -0.19942 0.0412 2982 -4.836  <.0001

TeD = TeD2:
 contrast        estimate     SE    df t.ratio p.value
 No_TL - Yes_TL -0.03939 0.0412 2982 -0.955  0.3395

TeD = TeD3:
```

```
   contrast       estimate    SE   df t.ratio p.value
 No_TL - Yes_TL -0.11689 0.0412 2982 -2.835  0.0046

TeD = TeD4:
 contrast       estimate    SE   df t.ratio p.value
 No_TL - Yes_TL -0.00239 0.0412 2982 -0.058  0.9538

TeD = TeD5:
 contrast       estimate    SE   df t.ratio p.value
 No_TL - Yes_TL -0.01994 0.0412 2982 -0.484  0.6286

TeD = TeD6:
 contrast       estimate    SE   df t.ratio p.value
 No_TL - Yes_TL -0.27753 0.0412 2982 -6.730  <.0001

TeD = TeD7:
 contrast       estimate    SE   df t.ratio p.value
 No_TL - Yes_TL -0.17210 0.0412 2982 -4.174  <.0001
```

Finally we have a few more significant results. We see that there are significant differences(p.<0.001) for the following datasets: TeD1, TeD3, TeD6 and TeD7. We can see that in all cases, the model scores are higher for when transfer learning was used with the highest impact on the regression dataset TeD6(difference of .28) followed by the classification dataset TeD1(difference of .20), the regression dataset TeD7 (0.17) and lastly, the classification dataset TeD3 (.12).