Random Intercepts in Ordinal Neural Networks

Matan Bendak and Yuval Benjamini

matan.bendak@mail.huji.ac.il
yuval.benjamini@mail.huji.ac.il

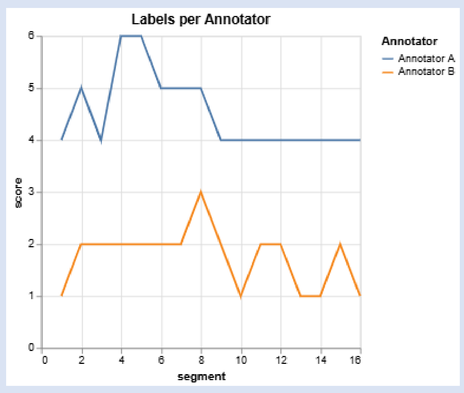
האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM
الجامعة العبرية في اورشليم القدس

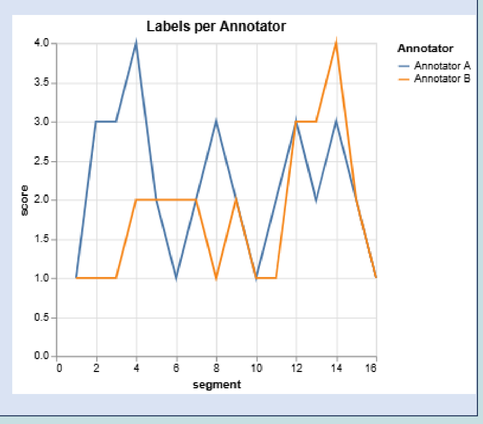
Statistics and Data Science,
The Hebrew University of Jerusalem

The data was collected by Itamar Jalon and Prof Talma Hendler

Motivation: Human annotated data of subjective

labels can have correlations between observations, these correlations can be addressed as different biases each annotator has.
We would like to suggest a way to remove the biases and to generalize to new, unobserved annotators.
Our data points are sentences of interviewees which were embedded using an LLM, each was human-annotated by psychology students. Those annotations are subjective as we can see in the figures below.





Ordinal Regression Neural Networks

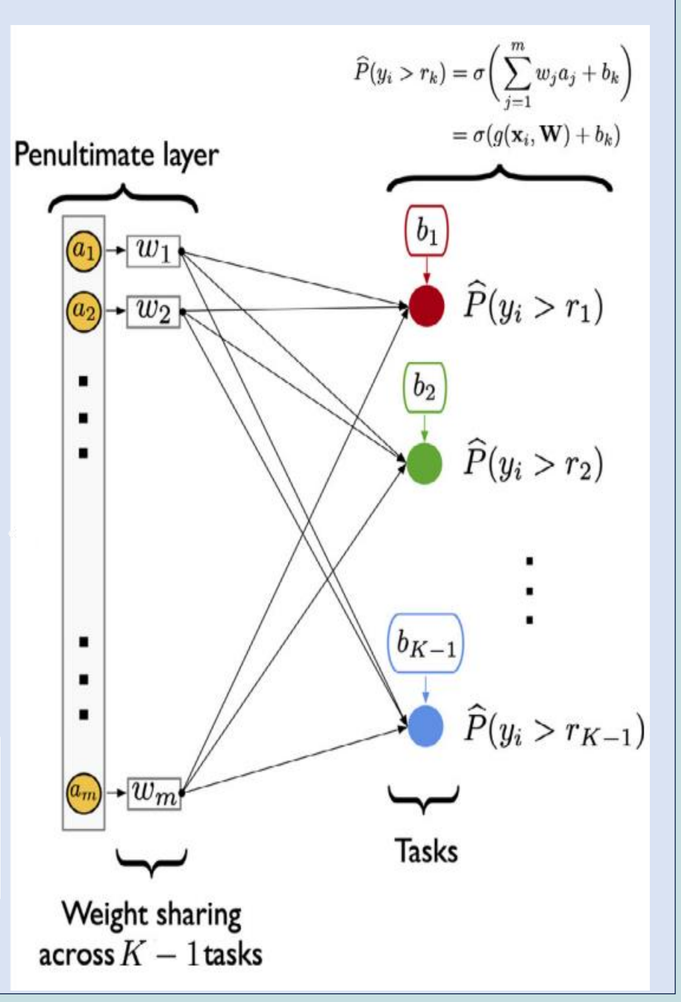
Defined as a multilabel task where each label k is the indicator "is the ordinal value larger than k?"

Consistency:
Can force consistency by adding non-shared Biases (CORAL).

Loss:

$$\mathcal{L} = \sum_{k=1}^{K-1} \mathrm{BCE}(y^{(k)}, \sigma(\theta - b_k))$$

$$\hat{P}(y_i > r_k) = \sigma\left(\sum_{j=1}^{m} w_j a_j + b_k\right)$$
$$= \sigma(g(\mathbf{x}_i, \mathbf{W}) + b_k)$$



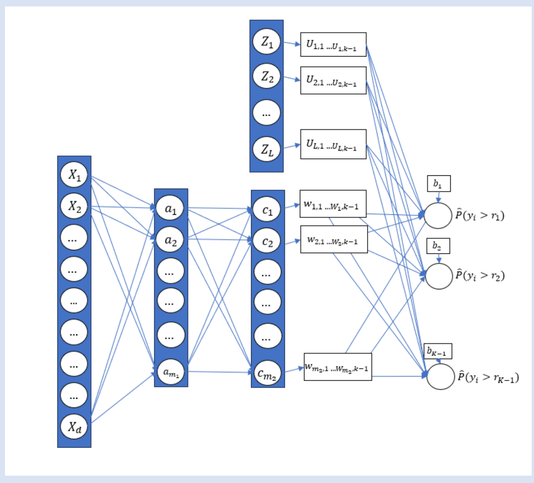
Random Intercepts:

Green Jr and Tukey [1960] explained when an effect is random, rather than fixed: "A model will be presented for these data that includes populations of P persons, T tests, and two halves. Then we can treat the experimental persons and the tests as samples from the populations of persons and tests. When a sample exhausts the population, the corresponding variable is fixed; when the sample is a small (i.e., negligible) part of the population, the corresponding variable is random".

$y_{\{i,j\}} = g(x_{i,j}; \theta) + b_j + \epsilon_{i,j}$
$y_{i,j}$: $outcome\ for\ observation\ i, by\ annotator\ j$
$g(x_{i,j}; \theta)$: $some\ function\ (NN\ for\ example)$
$b_j \sim N(0, \sigma_b^2)$: $random\ intercept\ for\ annotator\ j$
$\epsilon_{i,j} \sim N(0, \sigma^2)$: $noise$

Suggested Framework:

Integrate random intercepts into ordinal regression NNs by inserting a second input to the CORAL architecture.



Loss = Weighted Sum of Binary-Cross-Entropy over the Labels

Z = One Hot Encoding matrix of the Random Effects

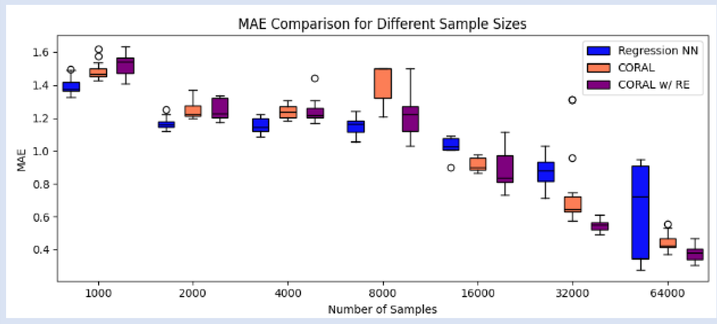
X = Numeric Features Matrix

W, U, b = Learnt weights and Biases of the model

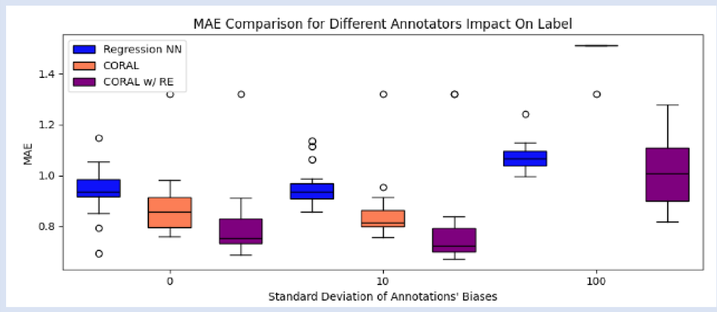
$$\hat{P}(y_i > k) = \sigma(g(x_i, Z_i\ U, W) + b_k)$$
$$= \sigma\left(h_x^{[1]}(U_k Z_k) + h_x^{[2]}(w_k, C_k) + b_k\right)$$

Simulation Results:

(MAE between the predicted and the real, ordinal values)



More data improves our model compared to the others



High biases' variance results in CORAL not converging.

Real Data Results: (MAE between predicted and annotated)

Emotion	Regression NN	CORAL	CORAL w/ RE
Irritation	1.43 (0.03)	1.32 (0.13)	**1.3** (0.03)
Nostalgia	1.38 (0.11)	1.29 (0.16)	**1.24** (0.04)
Pride	0.89 (0.04)	0.84 (0.0)	**0.8** (0.02)
Relief	0.77 (0.17)	0.67 (0.0)	**0.66** (0.01)
Sadness	1.3 (0.02)	1.18 (0.1)	**1.17** (0.05)
Satisfaction	1.02 (0.14)	1.05 (0.0)	**0.9** (0.02)
Surprise	0.77 (0.04)	**0.7** (0.0)	**0.7** (0.01)

Summary:

- Subjective annotations are frequently ignored in model development but should be addressed.
- We propose a method for integrating random intercepts into ordinal regression neural networks, which improves performance.