Assignment #2 – Uncertainty Prediction

https://github.com/Lemorita95/1FA006/tree/main/2_uncertainty

Written summary

Also available in https://github.com/Lemorita95/1FA006/blob/main/2_uncertainty/README.md

approach

1. hyperparameters:

   batch_size = 32

   learning_rate = 2e-4

   num_epochs = 10

2. Load data and stored in a CustomDataset class;
3. Data is normalized within CustomDataset class _*init*_ through z-score normalization;
4. Train, validation and test dataset is created with PyTorch Dataloader;
5. Training loss is defined as Negative Loss-likelihood at nll_loss();
6. Gives the user the possibility to load a model (if found at models/model.pth);
7. Default model is defined with 5 convolution layers (ReLU activation), 2 hidden layers (Softmax activation) and 1 output layers;
8. Output layer with 6 neurons (2 for each label), 3 for mean and 3 for log(standard deviation);
9. Train and validate model train_validade_model();
10. Test model test_model();
11. Plot pull histogram and scatter of averages;

results

1. the training-validation loop for the hyperparameters took around 2.2 minutes to complete;
2. training loss becomes negative in the 2nd epoch because of log(standard deviation) and is negatively larger then 0.5*log(2*pi);
3. training loss has a monotonic decrease throughout the epochs.
4. validation losses decreases throughout epochs.

challenges

1. training model are giving negative loss for some activation function;
2. when scatter plotting predicted mean vs true values with some activation functions (tanh, leakyrelu) some points with zero correlation can be seen, they also have high standard deviation (~0.98);
3. when using tanh as activation function for the last layer, pull histogram seems better but model predicts always predicts the same mean value - e.g. log(sigma) being more relevant on the loss function;
4. using percentile normalization also does not improved model prediction of mean and standard deviation, labels have higher bias;
5. softplus is the better so far;
6. softmax prevent negative losses but predicts discrete mean values;
7. lower learning rate prevent model for making a very distance first batch guess;
8. adding l2 regularization prevent model for not learning the mean (guessing zero mean);
9. adding more channels on data increase the performance but the time spent on training increased a lot;
10. adding more complexity to model increase its quality;
11. final configuration considered a good enough quality and time of execution of training;
12. the final model configuration overfitts around after 15 epochs;

Result plots

Figure 1: Train and validation loss per epoch.

From Figure 1 we see a decrease as the model goes through the epochs. This experiment used $Batch\ size = 32, Learning\ Rate = 0.0002, Epochs = 10$. When using 15 epochs (not shown in the chart) the model starts to present overfitting as the train loss slightly decreases and the validation loss started to increase. This overfitting comportment after 15 epochs was also observed in the previous assignment using the CNN to predict only the label.

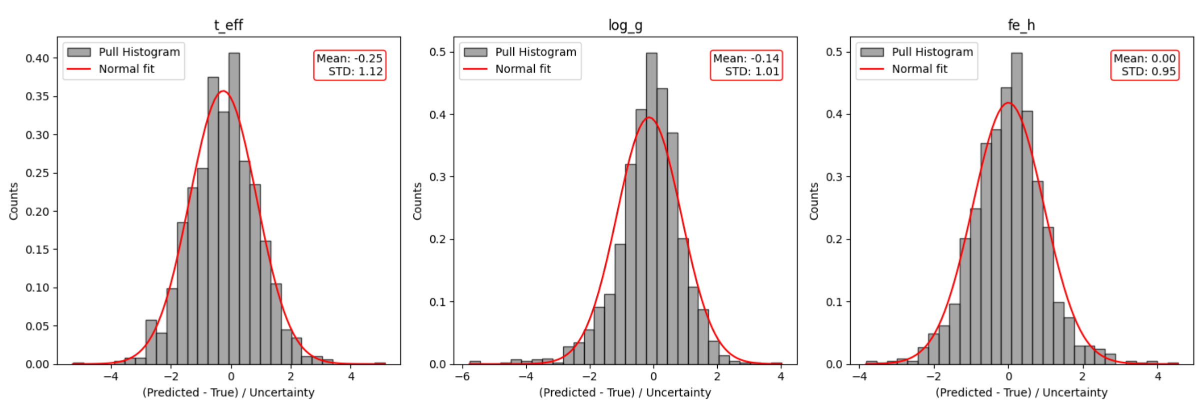

Figure 2: Pull histogram of predictions, distribution of residuals (predicted value – true value) normalized by the predicted standard deviation.

From Figure 2, we see the model performance varies for each label. Since the desired reference value is a normal distribution with zero mean and unit standard deviation, this CNN performs better for metallicity prediction then as for surface temperature and surface gravity, where bias can be observed. The model performs worst for surface temperature as in addition to bias, a higher-than-expected uncertainty is seen.

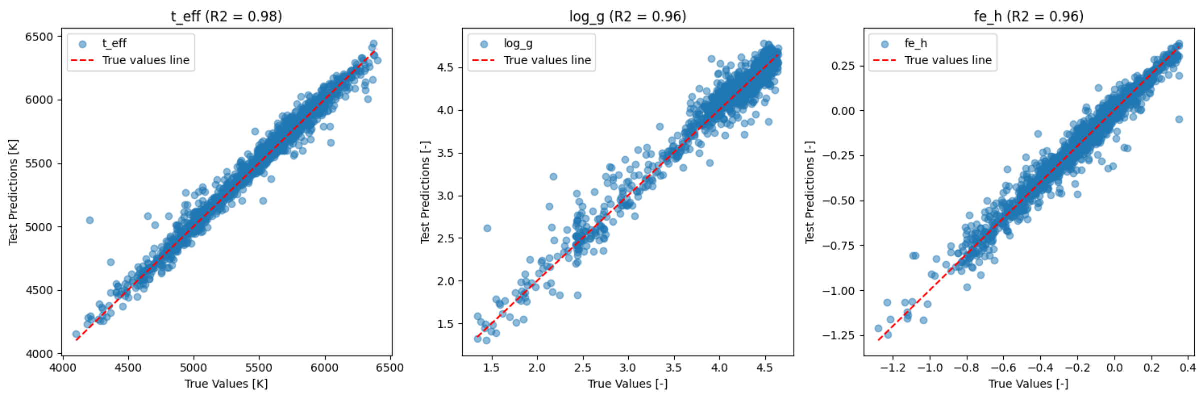

Figure 3: Scatter plot of predicted means.

Figure 3 firstly shows the limitation of using only R2 to evaluate a model performance. From those figures one might conclude a good performance model solely due to a close to 1 value of R2 those three labels. Although the predicted means presents good correlation to the true values, the predictions have some deviation from the expected mean and standard deviation as seen in Figure 2.

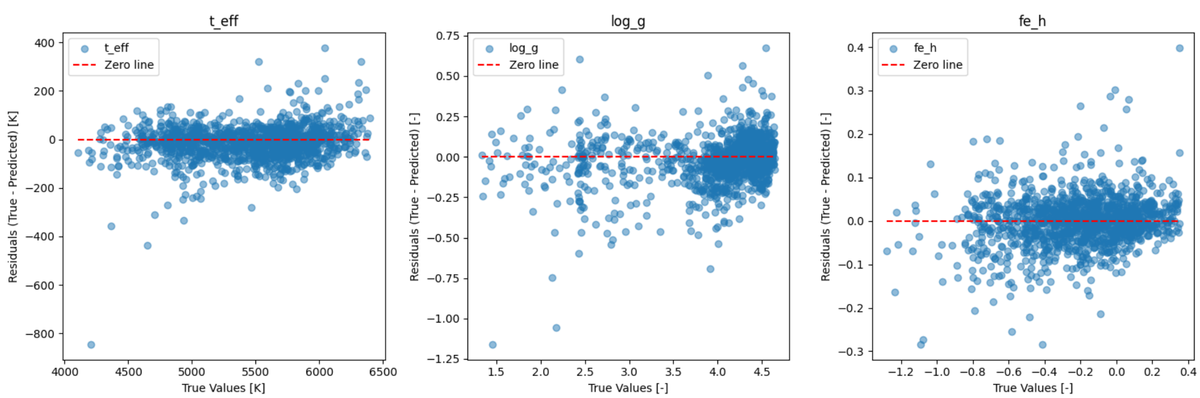

Figure 4: Scatter plot of predicted means residuals.

Figure 4 present the difference of the predicted means to the true values and large deviation can be seen for extreme values (minimum and maximum true value) for all three labels.