

NLP - ex2: Dror & Itay

(a) Which model had the highest accuracy?

The **Transformer model** achieved the highest accuracy. For instance:

- With 10% of the data, the Transformer reached a validation accuracy of **87.14%** by the third epoch.
- With 20% of the data, the Transformer reached a validation accuracy of **87.07%** by the third epoch.

In comparison:

- The **MLP with a hidden layer** achieved approximately **85.15%** with 100% of the data.
- The **Log-linear model** reached **85.28%** accuracy with 100% of the data.

(b) Which model was the most sensitive to the size of the training set?

The **Transformer model** showed the most sensitivity to training set size. With just 10% of the data, it achieved **87.14%**, closely matching its performance with 20% of the data (**87.07%**). This indicates its ability to generalize well even with limited data.

In contrast:

- The **Log-linear model** demonstrated a gradual improvement in accuracy as the training portion increased, highlighting its reliance on larger datasets.
- The **MLP with a hidden layer** also improved progressively with more data, though it plateaued earlier compared to the log-linear model.

(c) Extract and report the number of trainable parameters. Do additional parameters help? Explain.

1. **Log-linear model parameters: 8004**
2. **MLP with hidden layer parameters: 1,002,504**
3. **Transformer model parameters: 82,121,476**

Based on the architecture:

1. **Log-linear model:** Only one layer is trainable, so it will have fewer parameters (e.g., $\sim 2000 \times 4$ for 2000 TF-IDF features and 4 classes).
2. **MLP with hidden layer:** Includes an additional hidden layer with 500 neurons, significantly increasing the parameter count compared to the log-linear model.

3. **Transformer model:** Has millions of trainable parameters, as it uses a pre-trained `distil roberta-base` model with a classification head.

Do additional parameters help? Yes, additional parameters help, but only up to a point:

- The Transformer model's high accuracy demonstrates the power of pre-trained models and large parameter counts, especially on smaller datasets.
- However, simpler models like the log-linear classifier performed reasonably well, highlighting that complexity isn't always necessary for straightforward tasks with sufficient data.





