NLP – HW2

Ido Shapira Yoav Rabinovich  
319021044 208275735

1a)

1b)

1d)

Perplexity = 113.313

2a)

2b)

A diagram of mathematical equations

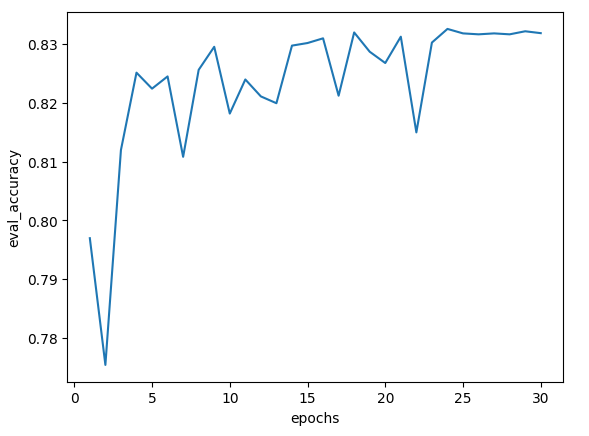
Description automatically generated with low confidence

3a)

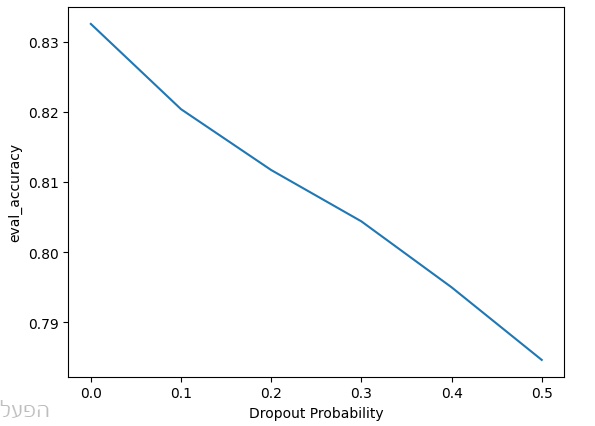
We think that character-based models will have better understanding of gentle differences between words and sentences. This will contribute to the persistence of “small window” elements of the language prediction like grammar. On the other hand, using character-based models with compared to word-based models will achieve less expressive results and won’t be able to understand semantic meaning as good as word-based models duo to the small granularity which emphasize gentle differences and not “big picture” ideas of sentences.

4.a.

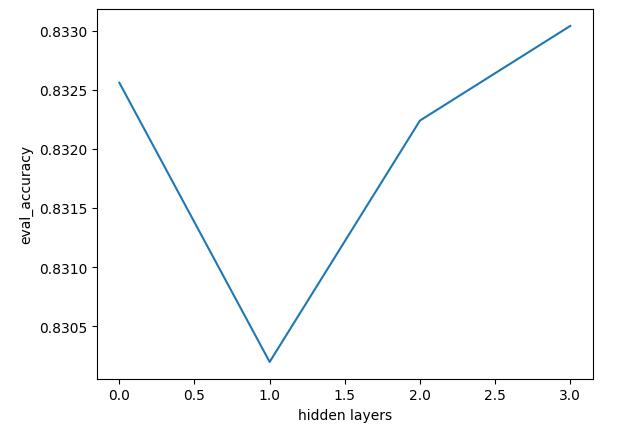
5. a.



b.



c.

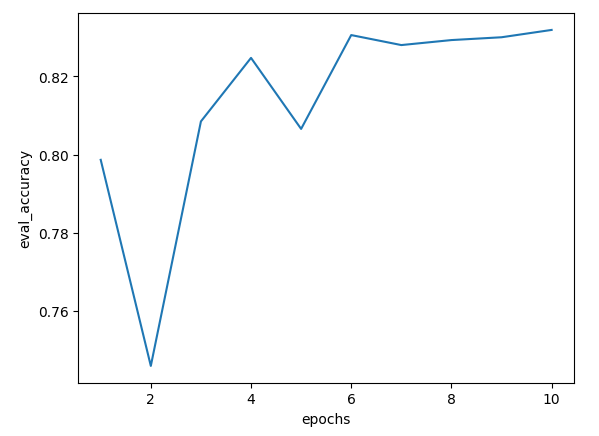


Before the experiment we expected to see better results for each added hidden layer. We didn’t expect that the model with one hidden layer will get worse results than with a linear model. We can see we still getting better results with 3 hidden layers but we expect that with more hidden layers we will reach the point where adding more hidden layers will have insignificant effect.

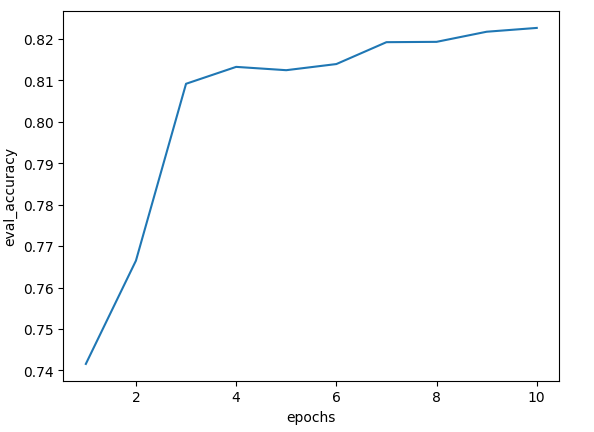
The linear models did not outperform the model with 3 hidden layers.

d.

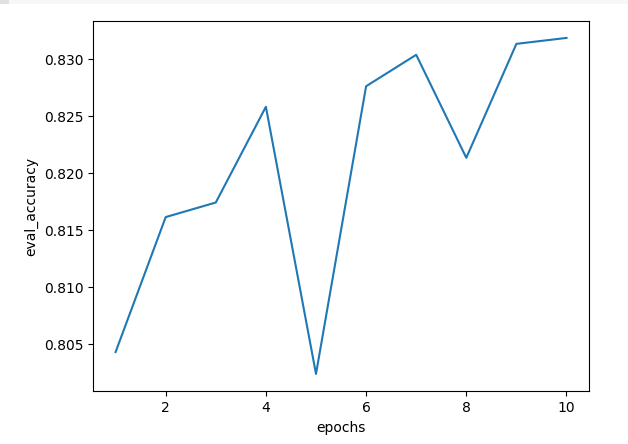
ReLU:



[Sigmoid:](https://pytorch.org/docs/stable/generated/torch.nn.Sigmoid.html#torch.nn.Sigmoid)



[Tanh](https://pytorch.org/docs/stable/generated/torch.nn.Tanh.html#torch.nn.Tanh):



We can see that there are differences between models with different activation functions. The best activation for this model is ReLU. We can see that the sigmoid activation function achieved close to peak results faster than the 2 other functions and was the steadiest among the activation functions.

6.

Using bayes rule: