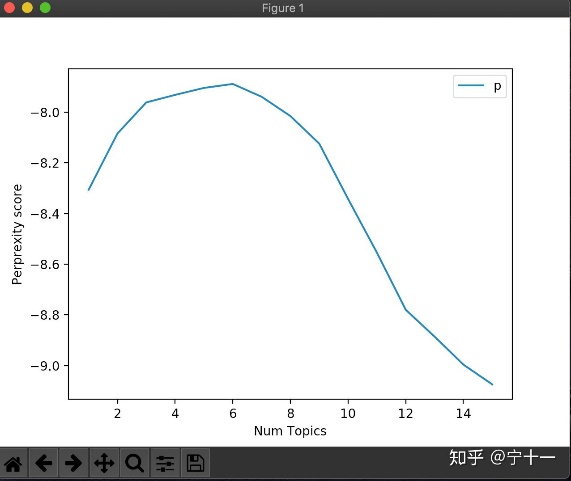
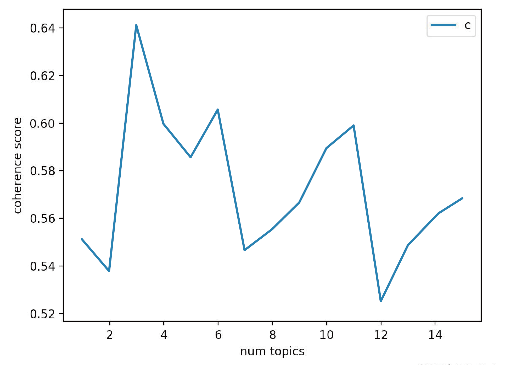
Finally, we used the LDA (Latent Dirichlet Allocation) topic model (LDA interpretation and literature) to classify the text of the topics related to the regulations of the four groups. Taking Children/Youth as an example, we have built 15 models in assignment 2. For which model is the most suitable topic model, we use perplexity and coherence to judge.

Perplexity is used to measure how well a probability distribution or probability model predicts the sample. It can also be used to compare the pros and cons of two probability distributions or probability models on the prediction samples. The probability distribution model or probability model with low perplexity can better predict the samples. (the perplexity result according to the Children/Youth text as shown in the figure)

From the above, we should choose the number of topics corresponding to the lowest point of perplexity. As shown in the chart, the more themes, the lower the score of perplexity, so we should choose 15 themes. But obviously this is wrong. Because when there are too many topics, our model has been overfitted. We found that when the number of topics exceeds 8, the perplexity of the model will continue to decrease. 8 is a turning point, so we consider the final number of topics from 1 to 7. As for how many topics we choose, we need to further use coherence to judge.