Forecast and Simulation **Summer Term 2023**
http://ml3.leuphana.de/lectures/summer23/FS/
Machine Learning Group, Leuphana University of Lüneburg

Ulf Brefeld (brefeld@leuphana.de)
Yannick Rudolph (yannick.rudolph@leuphana.de)

Exercise 6

Due: Thursday, June 8, 2023

Task 1: Conditional Random Fields for NER

This is a continuation of the NER task of Exercise 2.

Use the provided notebook *exercise-06-CRF.ipynb* to train a CRF model for NER. You will need to finish the implementation (look at the `#TODO` comments). Experiment with different observation features (see the `encode_sentence(...)` method) in order to get the best performance on the DEV set. Remember that you can use any input features (word, POS tag, chunk tag) from any token to define observation features for one token x_t. Observe that we use just one feature encoder (`x_enc`) for all observation features. It means that in case you use the same input feature (let us say *word*) from different tokens (x_t and the next token x_{t+1}) you should prefix the feature name with different strings (e.g. `word=xxx` and `word+1=xxx`), so that the model treats the features differently.

Answer

See *solution-06-CRF.ipynb*.

Task 2: Expectation Maximization

Please follow up the last lecture on Expectation Maximization (EM) by reading more thoroughly into the topic. We suggest reading through Chapter 9 of [Bishop, 2006]. Specifically you should understand the exposition in 9.3.3 on *Mixtures of Bernoulli distributions*.

Answer

Discussion in class.

Task 3: Mixtures of Bernoulli distributions

This Task builds upon the exposition in Section 9.3.3 of [Bishop, 2006].

i) Show that if we want to maximize the expected complete-data log likelihood function (Equation 9.55 in [Bishop, 2006]) for a mixture of Bernoulli distributions wrt. the mixing coefficients π_k (using a Lagrange multiplier to enforce the summation constraint), we obtain Equation 9.60.

ii) Reimplement the experiment described in Section 9.3.3 of [Bishop, 2006] (and illustrated in Figure 9.10). You will have to implement the EM algorithm for mixtures of Bernoulli distributions. We provide a fixed dataset in *exercise-06-EM.ipynb* (which involves downloading MNIST).

Answer

Task 3 pushed to next exercise sheet.

References

[Bishop, 2006] Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.