

CS 676

Final Project Appendix

Josh Wheeler, Yotam Barnoy

December 19, 2013

This document serves as an appendix to the presentation on the final project. Refer to that document for details.

1 Derivations of Gradient

Likelihood of datum m :

$$L = \exp\left(\sum_{t=2}^T \sum_{i=1}^F \lambda_i f_i(y_t^m, y_{t-1}^m, x^m)\right) / Z_{x^m}$$

Log Likelihood:

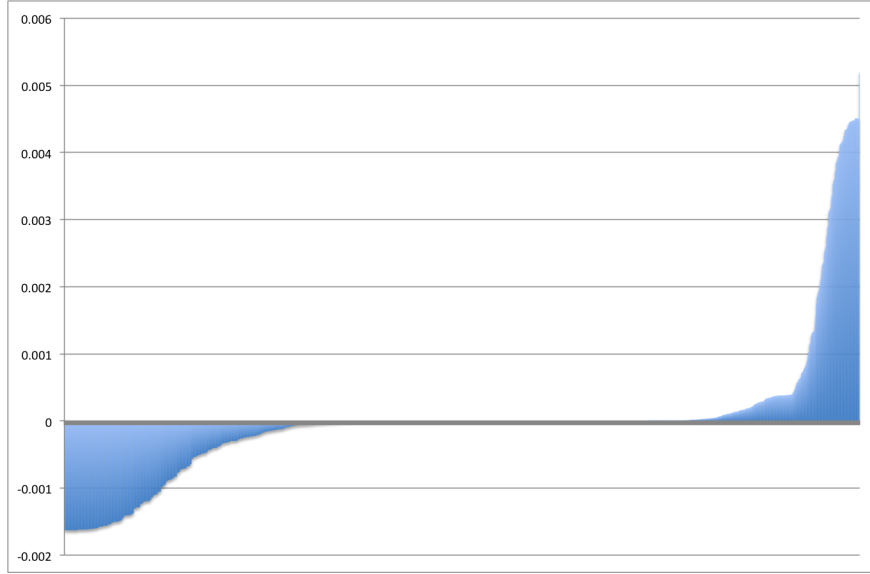
$$\begin{aligned} \text{LL} &= \sum_{t=2}^T \sum_{i=1}^F \lambda_i f_i(y_t^m, y_{t-1}^m, x^m) - \log Z_{x^m} \\ \frac{\partial \text{LL}}{\partial \lambda_i} &= \sum_{t=2}^T f_i(y_t^m, y_{t-1}^m, x^m) - \frac{\partial z}{\partial \lambda_i} Z \\ &= \sum_{t=2}^T f_i(y_t^m, y_{t-1}^m, x^m) - (1/Z) \sum_{y'} \exp\left(\sum_{t=2}^T \sum_{i=1}^F \lambda_i f_i(y_t'^m, y_{t-1}'^m, x^m)\right) \sum_{t=2}^T f_i(y_t'^m, y_{t-1}'^m, x^m) \\ &= \sum_{t=2}^T f_i(y_t^m, y_{t-1}^m, x^m) - \sum_{y'} (1/Z) \exp\left(\sum_{t=2}^T \sum_{i=1}^F \lambda_i f_i(y_t'^m, y_{t-1}'^m, x^m)\right) \sum_{t=2}^T f_i(y_t'^m, y_{t-1}'^m, x^m) \\ &= \sum_{t=2}^T f_i(y_t^m, y_{t-1}^m, x^m) - \sum_{y'} P(y'|x^m) \sum_{t=2}^T f_i(y_t'^m, y_{t-1}'^m, x^m) \end{aligned}$$

$$\begin{aligned}
&= \sum_{t=2}^T f_i(y_t^m, y_{t-1}^m, x^m) - \sum_{y'} \sum_{t=2}^T P(y'|x^m) f_i(y_t'^m, y_{t-1}'^m, x^m) \\
&= \sum_{t=2}^T f_i(y_t^m, y_{t-1}^m, x^m) - \sum_{t=2}^T \sum_{y'} P(y'|x^m) f_i(y_t'^m, y_{t-1}'^m, x^m) \\
&= \sum_{t=2}^T f_i(y_t^m, y_{t-1}^m, x^m) - \sum_{y'} P(y'|x^m) f_i(y_t'^m, y_{t-1}'^m, x^m)
\end{aligned}$$

Where the term on the right is expected value of f_i with respect to the current model parameters.

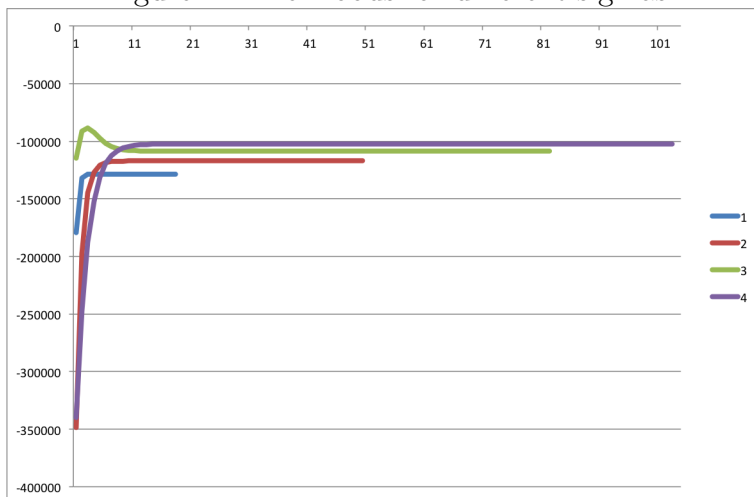
2 Graphs

Figure 1: Weights of feature functions



These graphs demonstrate some of our results. As can be seen in the first graph, our algorithm determines that some feature functions (and therefore some features) are far more correlated with some states than are others. Additionally, the second graph demonstrates the convergence of the likelihood for different values of sigma squared (1,

Figure 2: Likelihoods for different sigmas



2, 3 and 4 respectively) we see that convergence is slower with a higher sigma square, since the algorithm takes longer to get ‘trapped’ at a maximum point. However, the algorithm is often able to converge to a better likelihood given more time.

3 Sample data

Sample data for the features chi1, chi2 and hydrogen bonds can be found in the /data directory. After building the program with ‘build’, run ‘sample_run’ to see the program in action.

4 Result data

Please refer to the /results directory for some results we obtained using different parameters. Files called diffpart*.txt are taken from different batches of data, that is, not necessarily starting from the beginning of the data. Files called sigma*.txt vary the value of the sigma square parameter, and files called window*.txt vary the window size. We also ran some batch jobs with large amounts of data, and those are still running as we speak.