

Homework 2

Yawen Zhang

1. As shown in Fig.1, eta is set to be 0.001, 0.01, 0.1, 1 and 10, according to the convergence trend of HP, when eta is equal to 0.1, it has the fastest convergence speed (after finishing 4 passes, HP is near 0). When eta is too small, for example, 0.001 and 0.01, the convergence speed slows down, and the less the eta, the slower the convergence speed. When eta is too large, for example, 1 and 10, the trend of HP becomes very unstable and does not converge even after several passes.

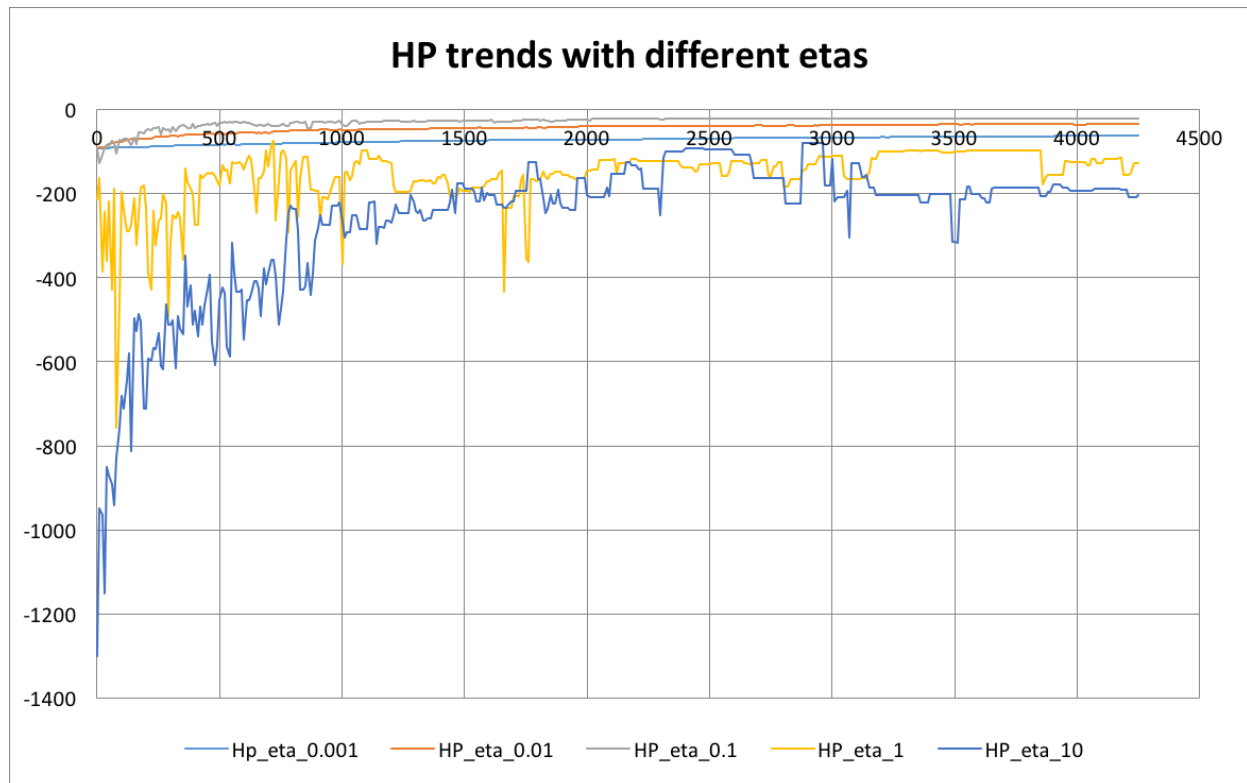


Fig.1 HP trend with different eta (0.001, 0.01, 0.1, 1, 10)

2. My stopping criterion is when HP stops increasing, that is, when the log-likelihood on the hold-out set stops increasing. Based on this criterion, I have tried "args.passes" from 1 to 10, and found that when it reaches 4 (or 5, their HP are very close) passes, HP becomes stable, thus, 4 passes over the data is needed to complete before stopping.

3. The best predictors for the positive class corresponding to the maximum w , while the predictors for the negative class corresponding to the minimum w . The reason for this is that the maximum absolute value of w makes the largest contribution to the probability. In conclusion, the best predictor for the positive class is 'baseball' while for negative class it is 'hockey'. I set $\eta = 0.1$ and $\lambda = 0.02$.

4. The poorest predictors of classes corresponding to w with absolute value equals to 0. When absolute w equals to 0, the corresponding x feature would have no contribution to the probability of each class. These predictors are 'aggravatingly', 'alway', 'baby', 'contained', 'crumbled', 'favour', 'pronunciation', 'prototypical', 'ratio', 'sportscasters', in total 10 poorest predictors.

(Note: Page 2 for extra problem)

Extra 1

As shown in Fig. 2, as the function is convex, two kinds of eta (whether with schedule or without schedule) both converge. However, there are still some difference between them, for eta without schedule, in some iterations, HP would drop to very low even when the iteration number reaches 3000, which indicates that eta may be too large in these iterations so HP decreases. For eta with schedule, this problem doesn't happen when iteration number reaches 3000, HP become relatively stable.

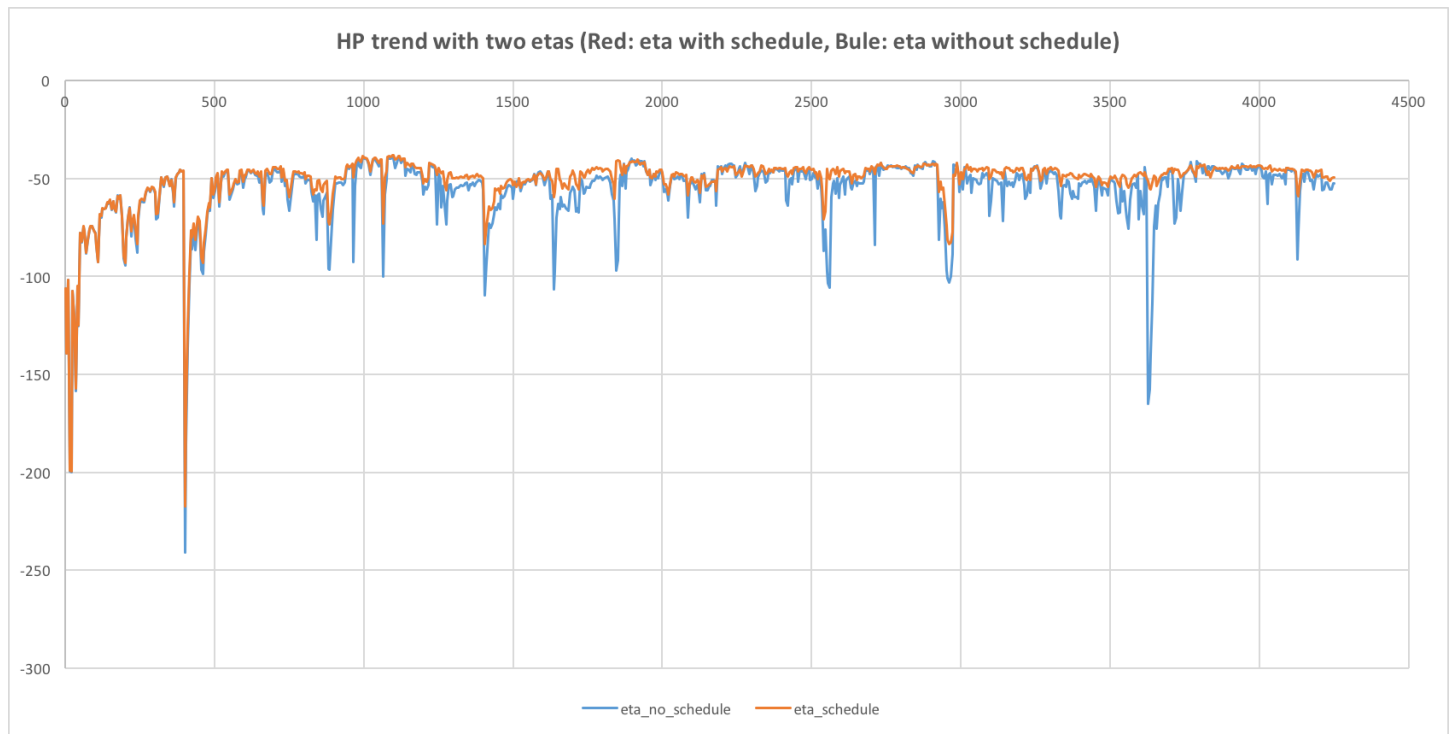


Fig.2 HP trend with two kinds of eta, one with schedule, the other without schedule