Lab II

Part (a)

The posterior probability can be computed using Bayes' theorem:

$$p_{X|Y_0^{N-1}}(x|y_0^{N-1}) = \frac{p_X(x)p_{Y_0^{N-1}|X}(y_0^{N-1}|x)}{p_{Y_0^{N-1}}(y_0^{N-1})}$$

Given the problem's assumption that the user ratings are independent and identically distributed, the likelihood can be factored as:

$$p_{Y_0^{N-1}|X}(y_0^{N-1}|x) = \prod_{n=0}^{N-1} p_{Y_n|X}(y_n|x)$$

where $p_{Y_n|X}(y_n|x)$ is the probability that user n gives rating y_n given the movie quality x.

Substituting this into the Bayes' theorem formula gives:

$$p_{X|Y_0^{N-1}}(x|y_0^{N-1}) = \frac{p_X(x)\prod_{n=0}^{N-1} p_{Y_n|X}(y_n|x)}{p_{Y_0^{N-1}}(y_0^{N-1})}$$

$p_{Y_0^{N-1}}(y_0^{N-1})$ ensures the posterior sums to 1:

$$p_{Y_0^{N-1}}(y_0^{N-1}) = \sum_{x' \in \mathscr{X}} p_X(x') \prod_{n=0}^{N-1} p_{Y_n|X}(y_n|x')$$

The posterior probability distribution for the true movie quality x, given the user ratings y_0^{N-1}, is:

$$p_{X|Y_0^{N-1}}(x|y_0^{N-1}) = \frac{p_X(x)\prod_{n=0}^{N-1} p_{Y_n|X}(y_n|x)}{\sum_{x' \in \mathscr{X}} p_X(x')\prod_{n=0}^{N-1} p_{Y_n|X}(y_n|x')}$$

Part (e)

I would recommend 'Schindler's List (1993)', because it is one of the movies that have the highest MAP rating of 10.0. The movie you should avoid is 'Wild Wild West (1999)', because it has the lowest MAP rating, which is 4.0.

Part (h)

From the result I got, we can observe that as the number of observations increases, the entropy decreases. As N approaches around 150-200, the entropy approaches zero, which means that additional data doesn't significantly reduce uncertainty further.

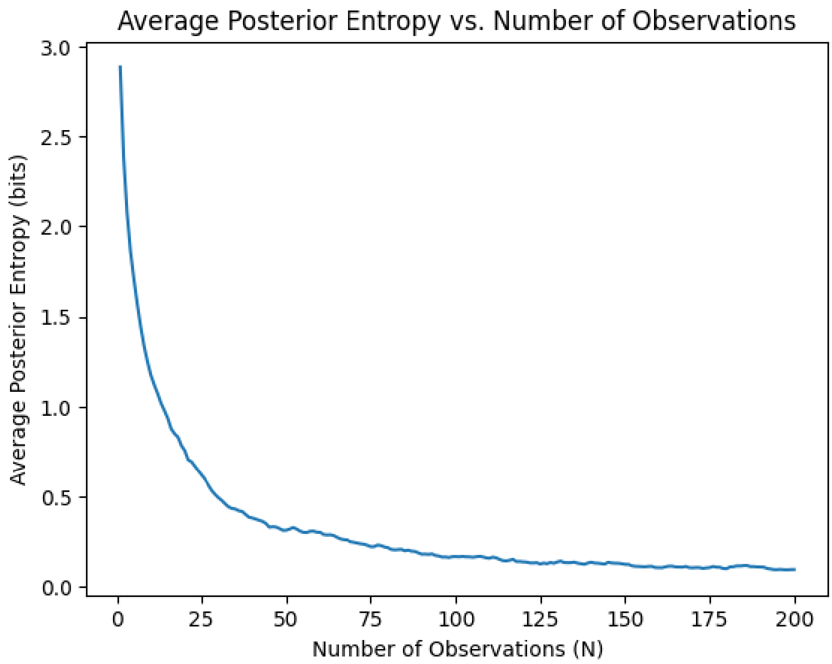

Figure 1: Average Entropy of the Posterior Distribution of the Movie Quality