IN-STK5000 Project 2 Report

...

Fall 2020

1 Historical data

write here

2 Improved policies

2.1 Exercise 2 Improved policies

For the improved policy we are doing exacly as suggested, "simply selecting, for each x_t, the action maximising expected reward according to your model", and for our model we are using logistic regression conditioned on a. There

2.1.1 Expected utility for the improved policy $\hat{\pi}$

The expected utility, given the improved policy $\hat{\pi}$ and model parameters θ, trained on the historical data, can be written as:

$$\begin{aligned} E_{\theta}^{\hat{\pi}}[U] &= E_{\theta}^{\hat{\pi}}[\sum_{t=1}^{T} r_t] = \sum_{t=1}^{T} E_{\theta}^{\hat{\pi}}[r_t|x] \\ E_{\theta}^{\hat{\pi}}[r_t] &= E_{\theta}^{\hat{\pi}}[y_t - 0.1a] \\ &= E_{\theta}^{\hat{\pi}}[y_t - 0.1a|a=0]p_{\theta}^{\hat{\pi}}(a=0) + E_{\theta}^{\hat{\pi}}[y_t - 0.1a|a=1]p_{\theta}^{\hat{\pi}}(a=1) \\ &= E_{\theta}^{\hat{\pi}}[y_t|a=0]p_{\theta}^{\hat{\pi}}(a=0) + \left(E_{\theta}^{\hat{\pi}}[y_t|a=1] - 0.1\right) p_{\theta}^{\hat{\pi}}(a=1) \end{aligned} \tag{1}$$

3 Adaptive experiment design

write here