4 Many-Armed Bandit and Early-Stopping

The problem of hyper-parameter optimisation can also be framed as a many armed bandit problem. A bandit problem describes a problem where an agent has a number of possible actions which yield a reward. Initially the association between an action and its reward is unknown to the agent. In a classic stochastic bandit problem, each action produces a reward randomly based on the probability distribution of reward linked to that specific action. The agent must therefore make a series of actions to gain information about reward associated with each action while also attempting to maximise the total reward. Figure 10 shows an example of this type of problem.

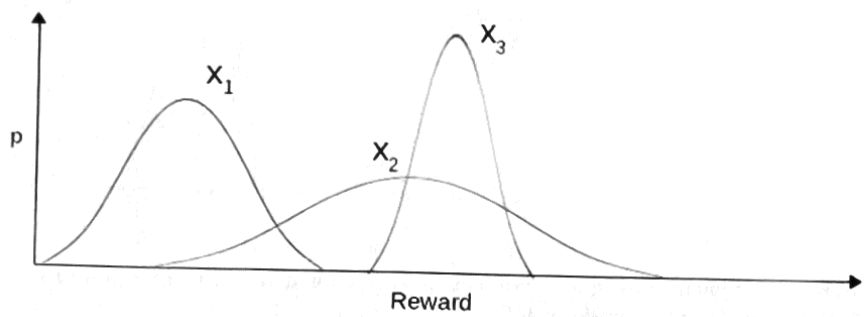

Figure 9: Bandit problem example with three actions denoted as x_1,x_2 and x_3. Each of these actions has an probability curve over a range of reward values which represents an agents belief about the value of each action. In this case there is a much higher uncertainty in x_2 than x_3 which has a higher mean value. A highly greedy approach to this problem would lead to the exploitation of x_3 rather than the exploration of x_2, which may not the optimal solution.

Hyper-parameter optimisation can be considered a non-stochastic variant of this problem. In this case an action or 'pull' becomes a single, or collection of, training iterations with a set of hyper-parameters. The models cost function or loss on the validation data set becomes the associated reward. In this application it is only the simple regret that is of interest rather than the cumulative regret as it is only the final recommendation which the search is evaluated on.

What is regret?

$$l_{\theta,t} = \frac{1}{|VAL|} \sum_{i \in VAL} loss(f_{\theta,t}(x_i), y_i) \tag{28}$$

Equation 28 describes the reward in this application, the cost function over