Large N limit of the knapsack problem

Solving the knapsack problem using the large N approximation in statistical physics

1 Introduction

We have a collection of N objects where object i has value v_i and weight w_i and a knapsack that can hold a maximum weight of W. We want to fill the knapsack with a collection of objects that have a total maximum value while remaining below the weight limit. How do we determine the collection of objects to include in the knapsack?

This question is known as the knapsack problem and there are many standard algorithms to solve it. In this work we showed how to use an analytical statistical physics formalism to devise a new algorithm that is valid for large N.

2 Main Result

We represent the knapsack problem a bit more formally. We define the collection of objects in the knapsack as $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ with $x_i = 1$ if object i is in the knapsack and $x_i = 0$ otherwise, and we define our weight and value vectors as $\mathbf{w} \equiv (w_1, w_2, \ldots, w_N)$ and $\mathbf{v} \equiv (v_1, v_2, \ldots, v_N)$, respectively. Then our objective is to find $\mathbf{x}$ that

$$\text{maximizes } \mathbf{v}\cdot\mathbf{x} \text{ subject to the constraint } \mathbf{w}\cdot\mathbf{x} \leq W\,, \tag{1}$$

where $\mathbf{a}\cdot\mathbf{b} \equiv a_1b_1 + a_2b_2 + \cdots + a_Nb_N$. For simplicity, we take v_i and w_i to be positive integers. Thus W is also an integer satisfying $W \geq N$. Letting $\mathbf{x}$ define the microstate of a statistical physics system at temperature $T = \beta^{-1}$, we find the partition function for this system can be written as

$$Z_N(\beta\mathbf{v}, \mathbf{w}, W) = \frac{1}{2\pi i}\oint_\Gamma \frac{dz}{z^{W+1}}\frac{1}{1-z}\prod_{k=1}^{N}\left(1 + z^{w_k}e^{\beta v_k}\right), \tag{2}$$

where Γ is a closed contour about the origin in the complex plane. By taking this partition function to the $N \gg 1$ limit, we can find an approximate expression for $\langle x_i\rangle$, the probability that object i is included in the collection. In the end, we obtain the following algorithm

1. Begin with the set of weights and values denoted by $\mathbf{w}$ and $\mathbf{v}$, respectively. Select the hyperparameter T for system temperature. Then use a numerical solver to solve the following for z_0:

$$0 = -W + \frac{z_0}{1-z_0} + \sum_{i=1}^{N}\frac{w_i e^{\beta v_i}}{z_0^{-w_i} + e^{\beta v_i}}. \tag{3}$$

2. With z_0 ompute $\langle x_\ell\rangle$ from

$$\langle x_\ell\rangle = \frac{e^{\beta v_\ell}}{z_0^{-w_\ell} + e^{\beta v_\ell}}. \tag{4}$$

3. Select a probability threshold p_{thres}. For each $\ell = 1, \ldots, N$ take

$$X_\ell = \begin{cases} 1 & \text{if } \langle x_\ell\rangle > p_{\text{thresh}}, \\ 0 & \text{otherwise.}\end{cases} \tag{5}$$

The vector $\mathbf{X} = (X_1, X_2, \ldots, X_N)$ represents the final selection of objects for the knapsack.