Application of Gibbs sampling in meal plan recommendation algorithm

Guo Zheng

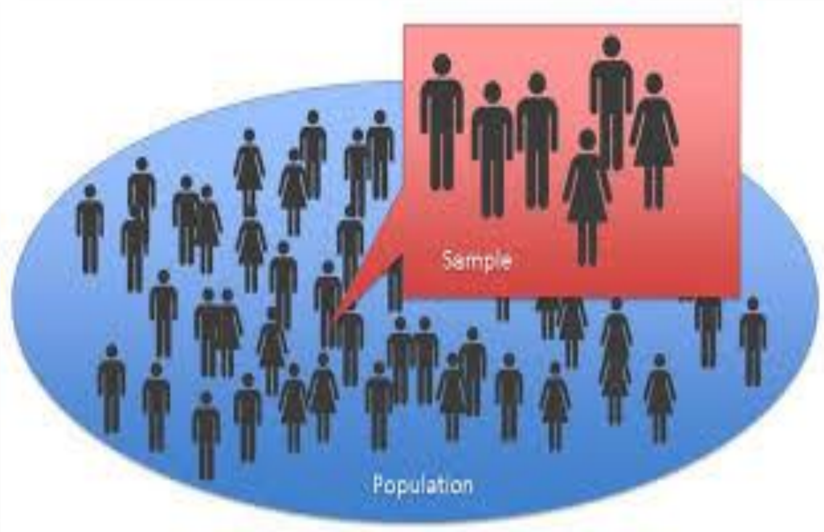
Sep 21, 2018

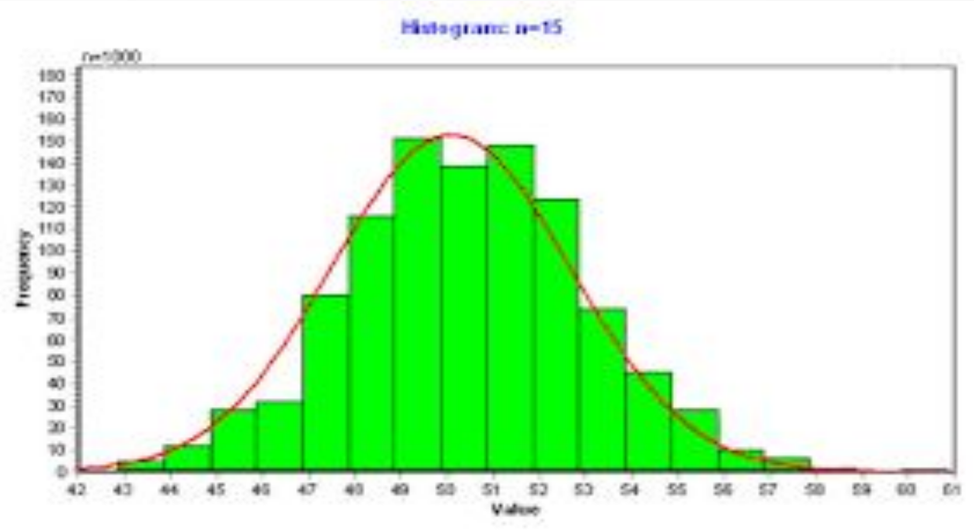
管 理 数 字 生 命 MANAGE THE DIGITALL LIFE

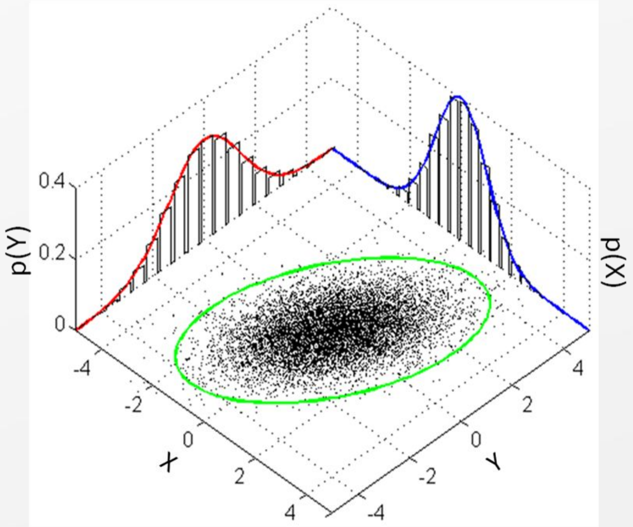
Outline

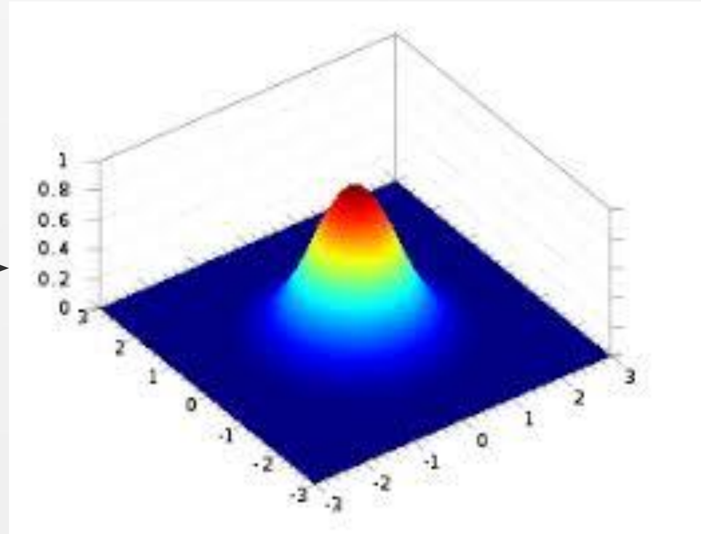
- AU
  - Gibbs Sampling Method (GS)
    - Idea of sampling
    - Gibbs Sampling overview
  - Meal plan recommendation algorithm
    - Application of GS
    - Other possible methods & their performances
    - Potential improvements

Idea of Sampling









ICX Research

Gibbs Sampling properties

- Sampling method for probability distribution
  - Multivariate
  - Monte Carlo Markov Chain process (MCMC)
    - Theoretical basis for convergence
  - Hard to sample directly

Gibbs Sampling pseudocode

- Algorithm

Algorithm 1 Gibbs sampler

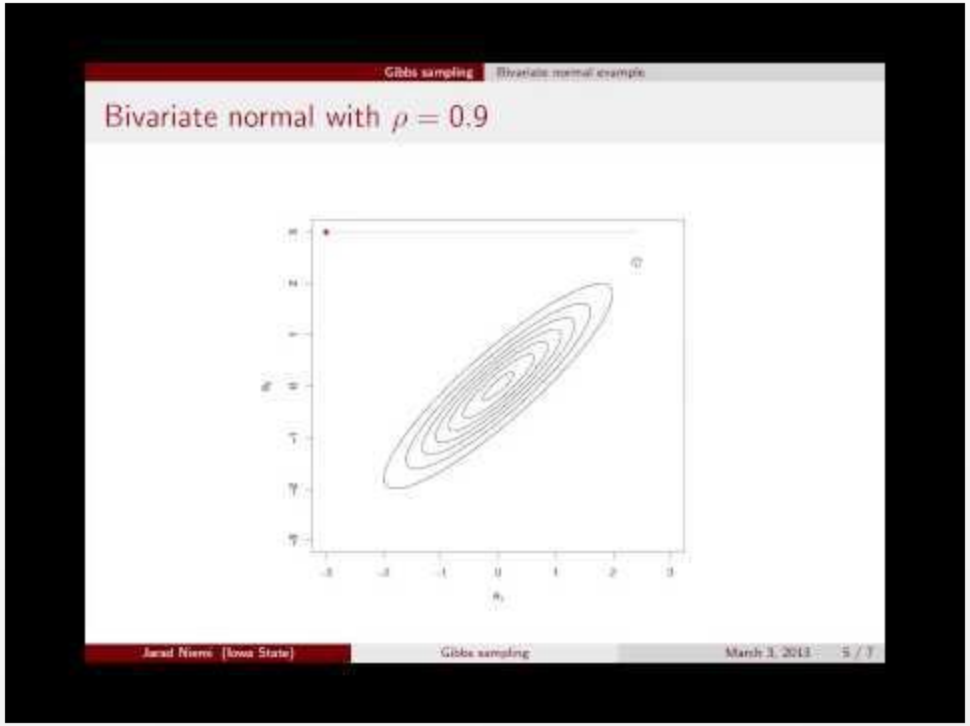
Initialize $x^{(0)} \sim q(x)$
for iteration $i = 1, 2, \dots$ **do**

$$x_1^{(i)} \sim p(X_1 = x_1 | X_2 = x_2^{(i-1)}, X_3 = x_3^{(i-1)}, \dots, X_D = x_D^{(i-1)})$$
$$x_2^{(i)} \sim p(X_2 = x_2 | X_1 = x_1^{(i)}, X_3 = x_3^{(i-1)}, \dots, X_D = x_D^{(i-1)})$$
$$\vdots$$
$$x_D^{(i)} \sim p(X_D = x_D | X_1 = x_1^{(i)}, X_2 = x_2^{(i)}, \dots, X_D = x_{D-1}^{(i)})$$

end for

Gibbs Sampling demonstration

- 5:09-5:49



ICX Research

Meal Plan Recommendation

- Context
  - Data:
    - About 1700 types of food divided into 13 categories (淀粉，蔬菜，etc)
    - Amount of nutrition components (35 components) for every food type
    - Human body’s requirements for every component (range)
  - Goal: combine a meal plan
  - Cost function
    - Number of requirements satisfied
  - Restrictions:
    - One type of food from one category

ICX Research

Mathematical expression

Constrained optimization problem

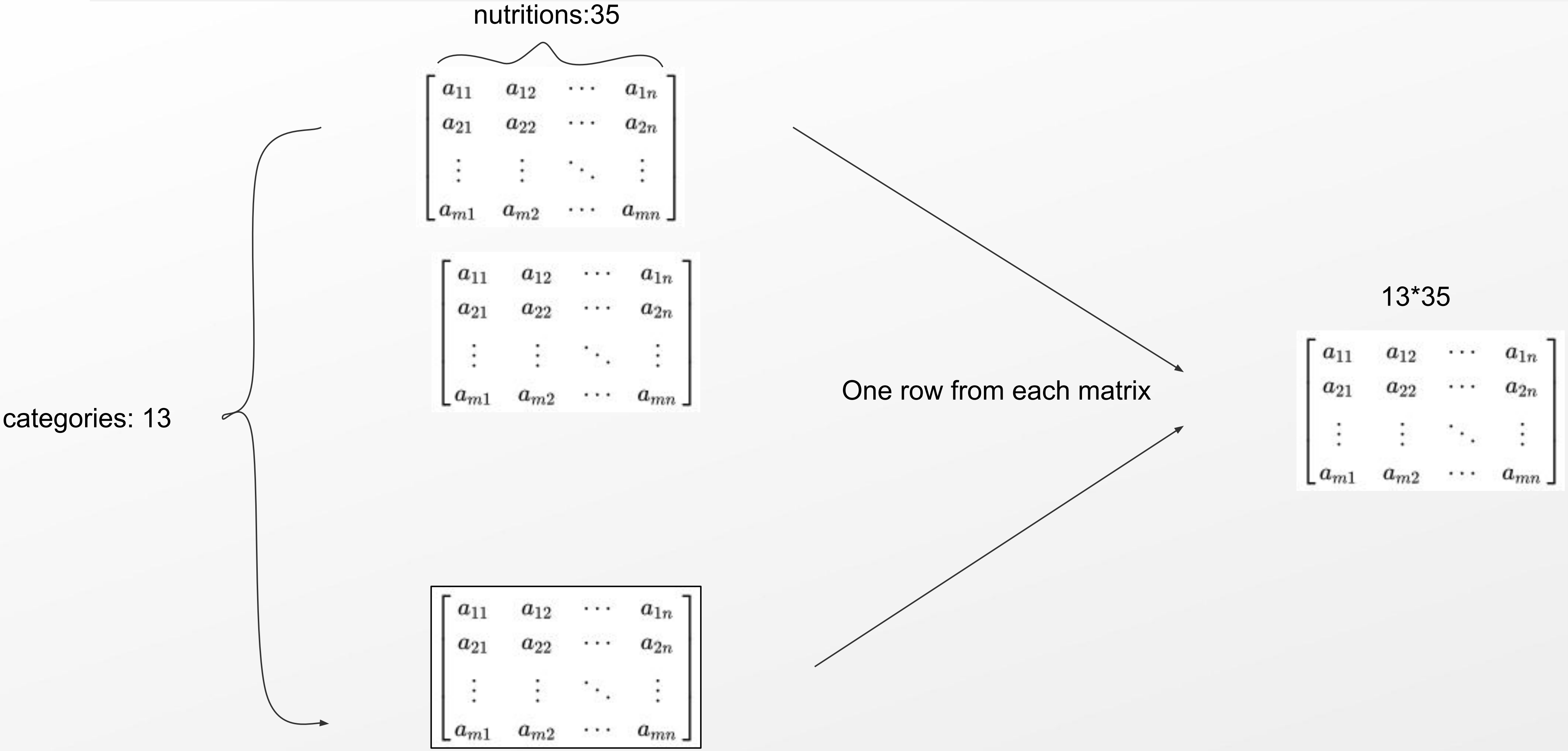
max: score(x)

subject to:

1. g(x) = 1
2. f(x) in range (min,max)

数字生命研究院
ICX Research

Meal Plan Recommendation



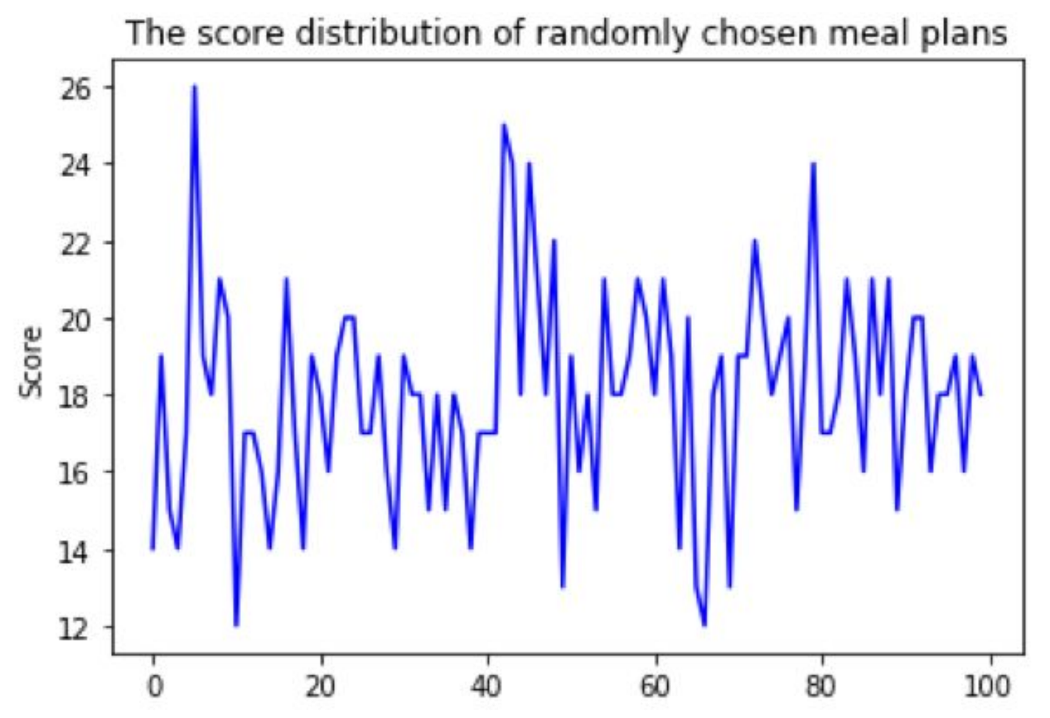
ICX Research

Meal Plan Recommendation

In total 9.66e+22 combinations!

Typically, 1 million trials gives about 10 meal plans scoring 30.

⇨ Brute force doesn’t work!



Meal Plan Recommendation

Gibbs Sampling ⇔ Meal Plan Recommendation

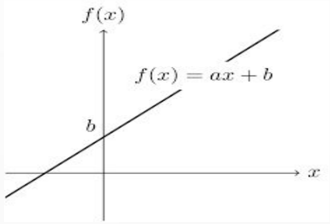
- Meal plan ⟺ 13-dimension probability distribution.

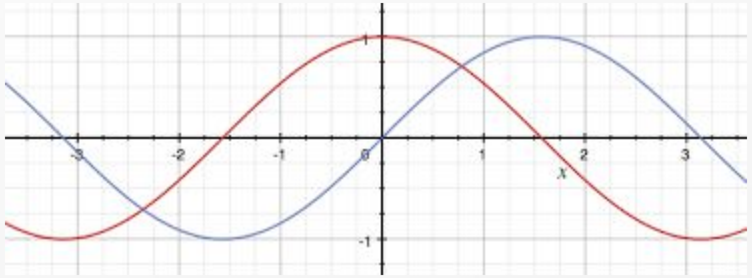
- Probability distribution determines the chance of being chosen to be the new initial point.
  - Relate score with probability. (Function choosing)

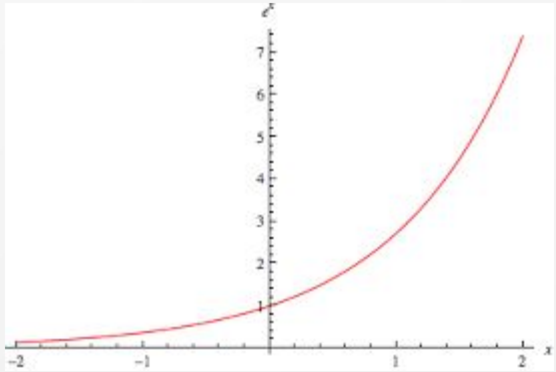
Mapping function

Choosing mapping function:

$$f(x) = \begin{cases} C * x \\ \sin(kx) \text{ (or cos, tan,etc)} \\ \exp(kx) \end{cases}$$



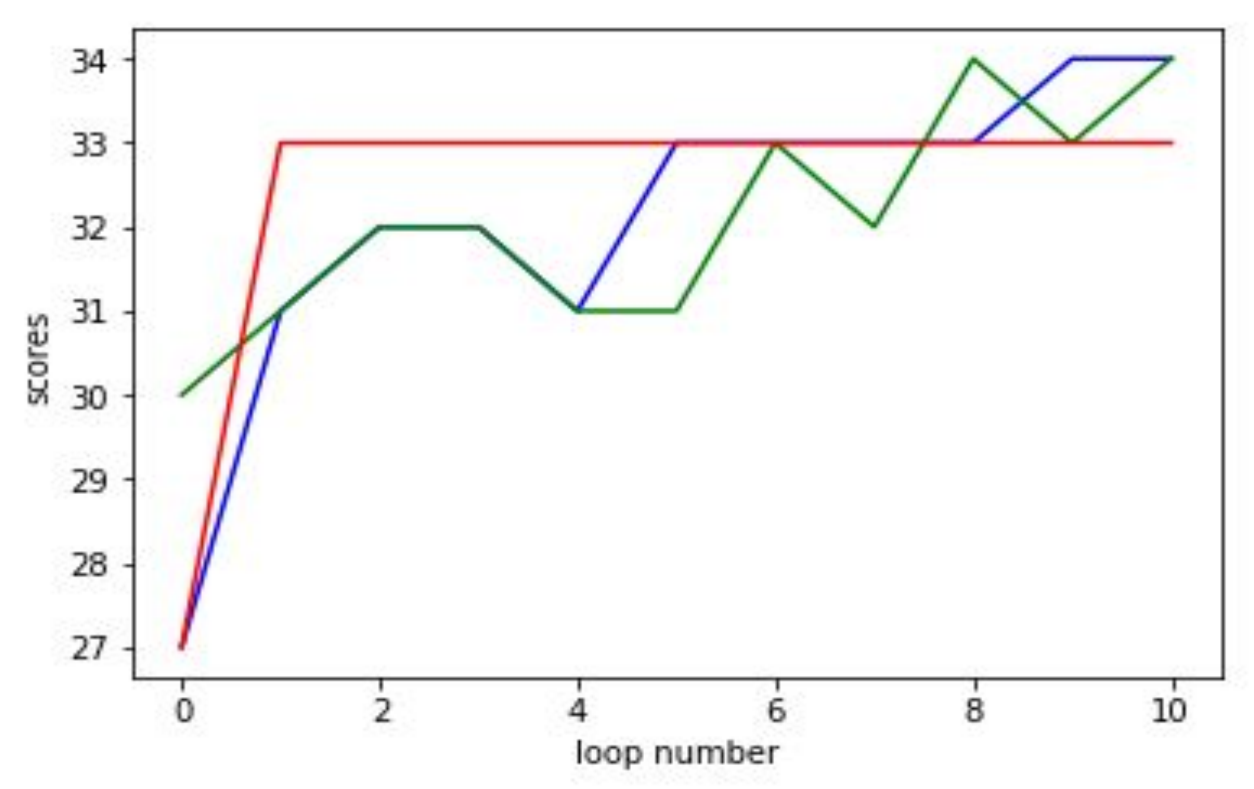




ICX Research

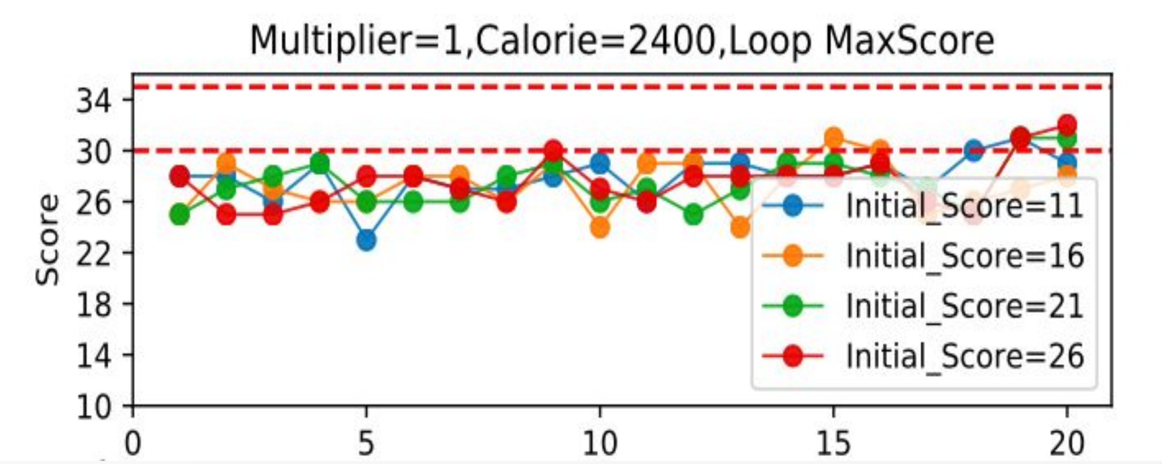
Meal Plan Recommendation

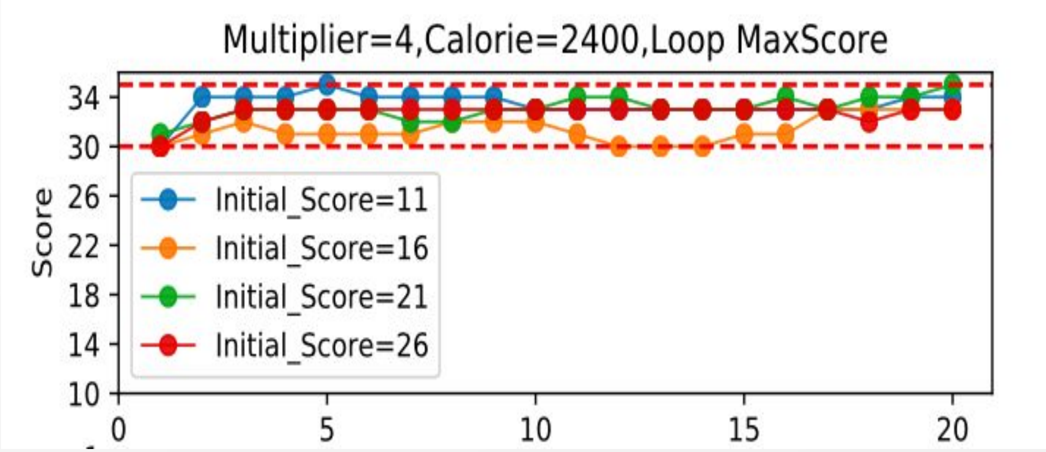
Comparison between gradient method and gibbs sampling:

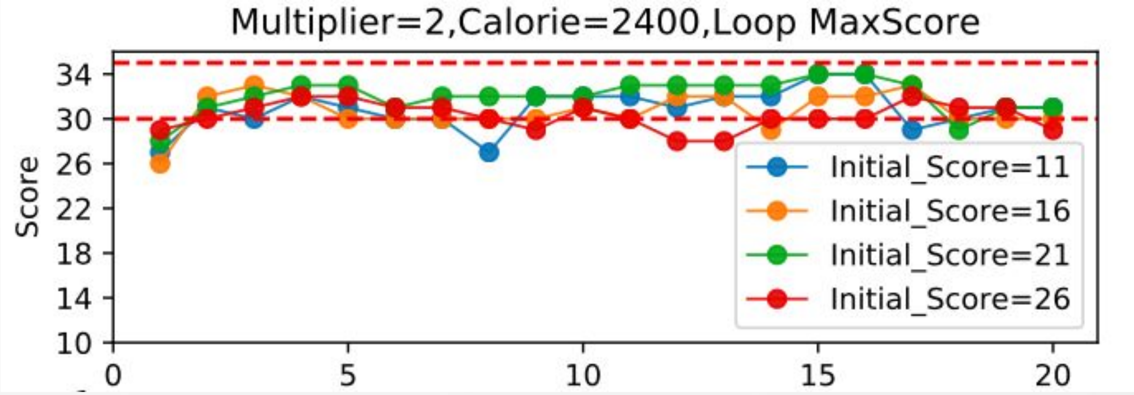


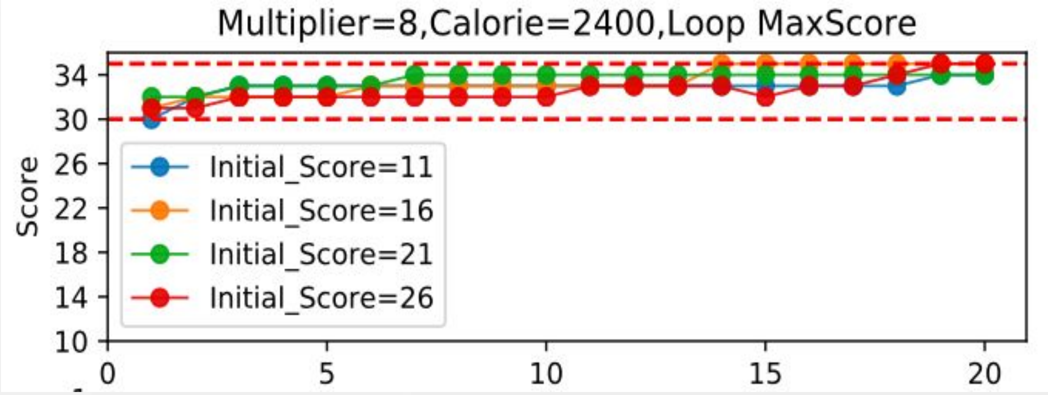
Multiplier coefficient examination

- Determination of the coefficient k: $f(x) = \exp(kx)$









Performance

- Algorithm final performance (multiplier: 4)
  - Finds meal plan with score 35 in an average loop number around 30 loops, 15 minutes

Other methods

- Genetic algorithm:
  - Basic idea: generate better offsprings through mutation, crossover, and selection
  - Current performance: much slower than Gibbs
- Annealing algorithm:
  - Basic idea: reduce the randomness of the selecting process for Gibbs sampling

Potential future improvements

1. More detailed examination on the optimal multiplier value

2. Setting threshold for loop number

3. Examination of the performance of annealing algorithm

References

- http://www.mit.edu/~ilkery/papers/GibbsSampling.pdf

Thanks ~