Mining massive Datasets WS 2017/18

Problem Set 7

Rudolf Chrispens, Marvin Klaus, Daniela Schacherer

December 9, 2017

Exercise 01

- Internally an array is used to store both - key and value - as an object of the so-called Entry class. For handling collisions meaning values that map to the same key a linked list is further used.

- A new key-value pair is added by the put-method. This method calculates the hash value for a given key and uses it as index in the hash table. If the calculated index position already holds a key-value pair, the linked list at this position is used and the key-value pair is stored at the first available position in this list. The function is also responsible for resizing the table if appropriate.

- When one wants to search given a key the corresponding index is calculated. At the calculated position the key is compared to the given search key and if it matches the corresponding value is returned. If it doesn't match, each entry of the linked list at this hash map position is searched through until the matching key is found. If the key does not exist this is also returned to the user.

Exercise 02

Exercise 03

If I understand that Question right I just should plot and look at whcih point the lines exceeds the value 0.5? If so then here the approximate values for each plot. I write the values for each r in a list that show the value of b as [10, 20, 50, 100].

r=2 - [0.26, 0.18, 0.13, -] (Figure 1)

r=3 - [0.4, 0.32, 0.23, 0.18] (Figure 2)

r=5 - [0.58, 0.5, 0.42, 0.34] (Figure 3)
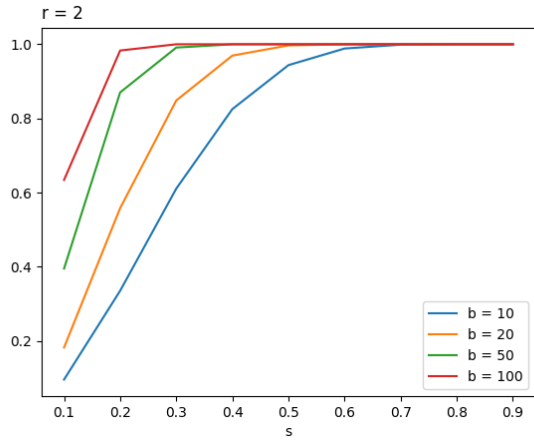
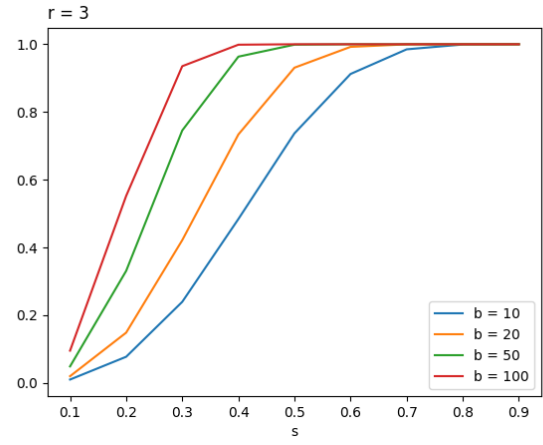r=7 - [0.67, 0.62, 0.53, 0.48] (Figure 4)
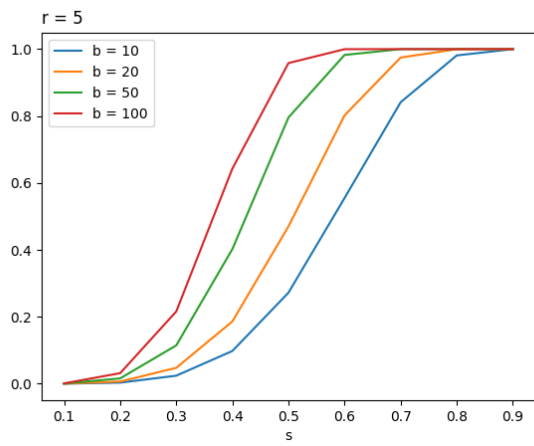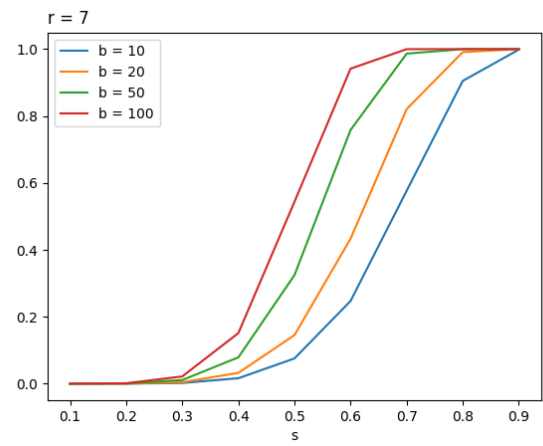
Figure 1 – r = 2

Figure 2 – r = 3

Figure 3 – r = 5

Figure 4 – r = 7

Exercise 04