

Intelligent Data Management - Exercise 3

Christoph Prinz

November 14, 2017

Assignment 1

Element	S_1	S_2	S_3	S_4	h_1	h_2	h_3
0	0	1	0	1	1	2	2
1	0	1	0	0	3	5	1
2	1	0	0	1	5	2	0
3	0	0	1	0	1	5	5
4	0	0	1	1	3	2	4
5	1	0	0	0	5	5	3

Minhash-Signature Step 1:

	S_1	S_2	S_3	S_4
h_1	∞	∞	∞	∞
h_2	∞	∞	∞	∞
h_3	∞	∞	∞	∞

Minhash-Signature Step 2:

	S_1	S_2	S_3	S_4
h_1	∞	1	∞	1
h_2	∞	2	∞	2
h_3	∞	2	∞	2

Minhash-Signature Step 3:

	S_1	S_2	S_3	S_4
h_1	∞	1	∞	1
h_2	∞	2	∞	2
h_3	∞	1	∞	2

Minhash-Signature Step 4:

	S_1	S_2	S_3	S_4
h_1	5	1	∞	1
h_2	2	2	∞	2
h_3	0	1	∞	0

Minhash-Signature Step 5:

	S_1	S_2	S_3	S_4
h_1	5	1	3	1
h_2	2	2	2	2
h_3	0	1	4	0

Final Minhash-Signature (After Step 6):

	S_1	S_2	S_3	S_4
h_1	5	1	3	1
h_2	2	2	2	2
h_3	0	1	4	0

Jaccard Similarities formula = $|S \cap T| \div |S \cup T|$

$$SIM(S_1, S_2) = SIM(\{2, 5\}, \{0, 1\}) = 0$$

$$SIM(S_1, S_3) = SIM(\{2, 5\}, \{3, 4\}) = 0$$

$$SIM(S_1, S_4) = SIM(\{2, 5\}, \{0, 2, 4\}) = \frac{1}{4}$$

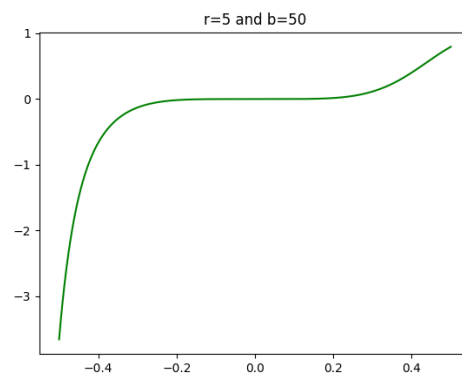
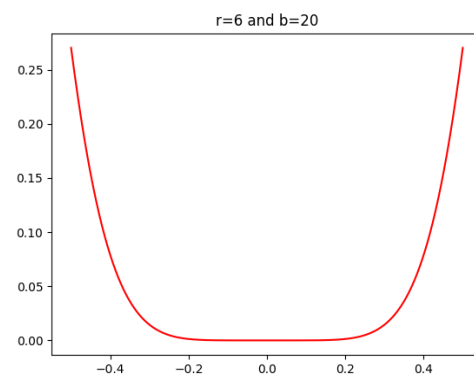
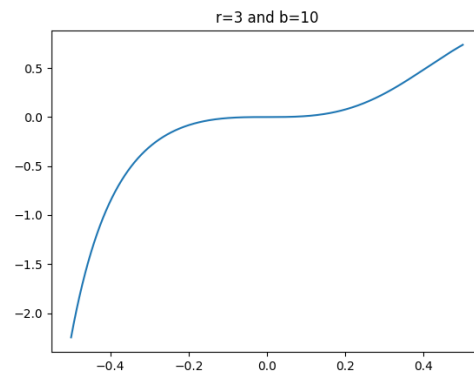
$$SIM(S_2, S_3) = SIM(\{0, 1\}, \{3, 4\}) = 0$$

$$SIM(S_2, S_4) = SIM(\{0, 1\}, \{0, 2, 4\}) = \frac{1}{4}$$

$$SIM(S_3, S_4) = SIM(\{3, 4\}, \{0, 2, 4\}) = \frac{1}{4}$$

Assignment 2

Plots for s-curve:



s that fulfils $1 - 1(1 - s^r)^b = \frac{1}{2}$:

a) $s_1 = \sqrt[3]{1 - \frac{1}{10\sqrt{2}}} \approx 0.406$

$s_2 = \sqrt[3]{1 + \frac{1}{10\sqrt{2}}} \approx 1.246$

$1/b^{1/r} \approx 0.464$

$$\text{b) } s_1 = \sqrt[6]{1 - \frac{1}{20\sqrt{2}}} \approx 0.569$$

$$s_2 = \sqrt[6]{1 + \frac{1}{20\sqrt{2}}} \approx 1.194$$

$$1/b^{1/r} \approx 0.607$$

$$\text{c) } s_1 = \sqrt[5]{1 - \frac{1}{50\sqrt{2}}} \approx 0.424$$

$$s_2 = \sqrt[5]{1 + \frac{1}{50\sqrt{2}}} \approx 1.147$$

$$1/b^{1/r} \approx 0.457$$