
DSC 190 - Discussion 08

Problem 1.

- a) Given a collection of 200 billion unique objects, design a Bloom filter with a 5% false positive rate by choosing the size c of the bit array and the number of hash functions, k .
- b) How much memory will your Bloom filter require, approximately?
- c) Suppose your machine has 32 gigabytes of memory. What false positive rate must you accept in order to fit the Bloom filter in memory?
- d) A Bloom filter using 32 gigabytes of memory has a false-positive rate similar to flipping a coin. In what ways is the Bloom filter better than a coin?