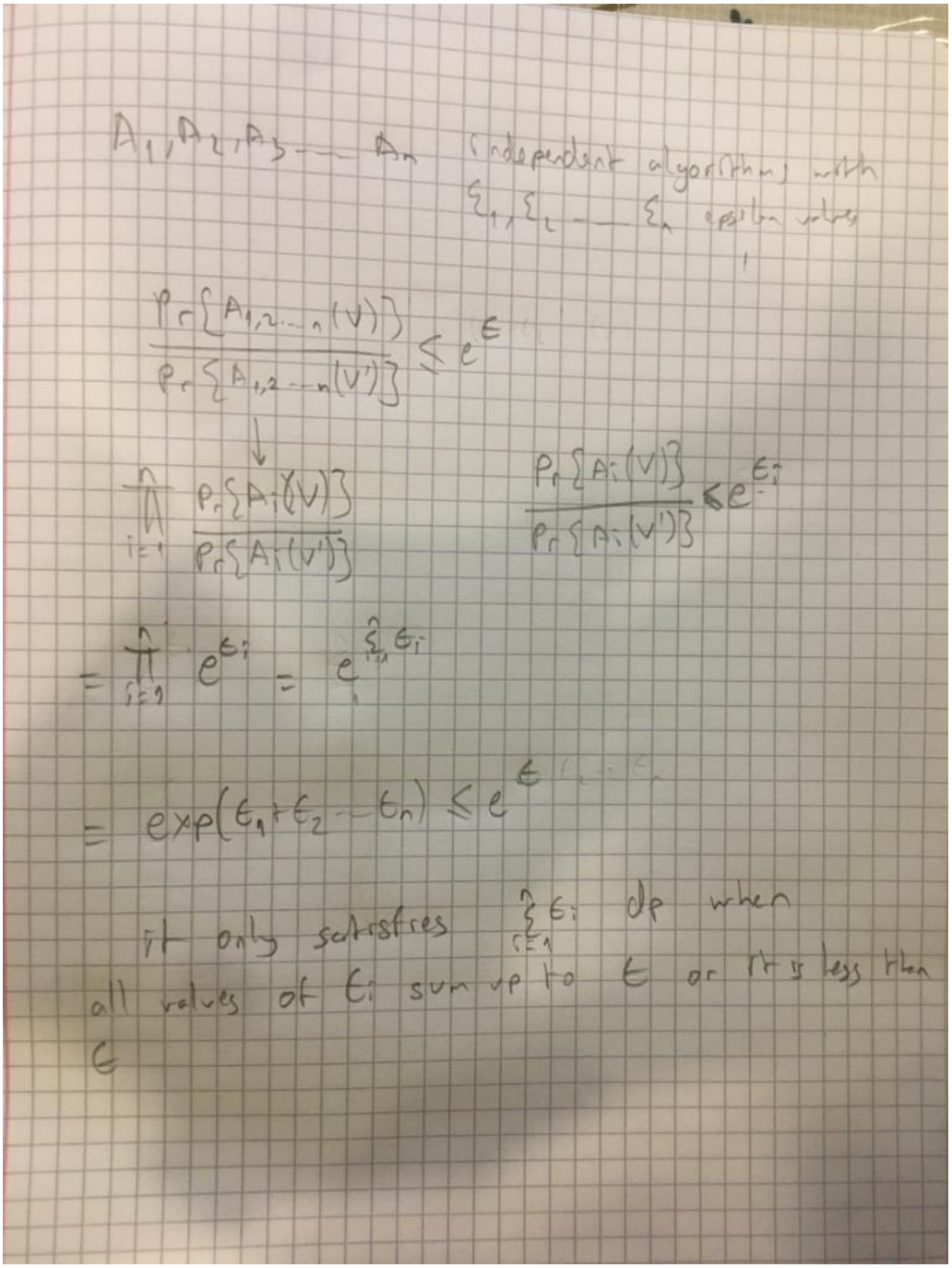
Ertan Can Güner
60610

COMP430 – Homework 2
Report

Part 1

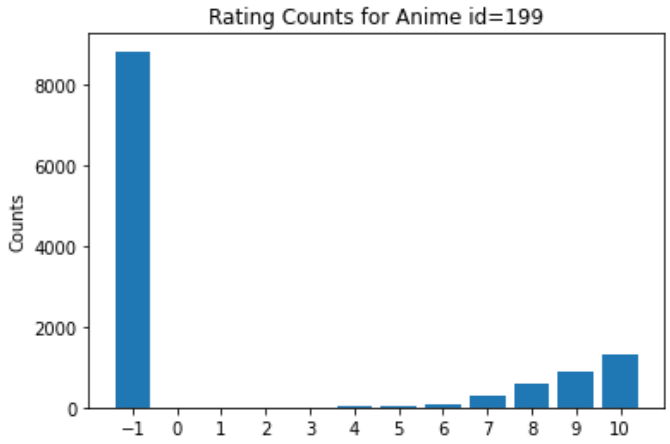


1. It does not satisfy sum(Ei)-DP because it is not always satisfied.

2. This algorithm does not satisfy E-DP because an adversary can query the algorithm and determine the pattern behind the algorithm and gain valuable information about the dataset.
3. Scale parameter in Laplace Noise is defined by b <= 2S(q)/E. In this question we use a scale parameter of b = E, which means that sensitivity of the dataset is 0.5 * E^2 which is troubling because a query is leaking information about the dataset itself so it does not satisfy E-DP.

Part 2

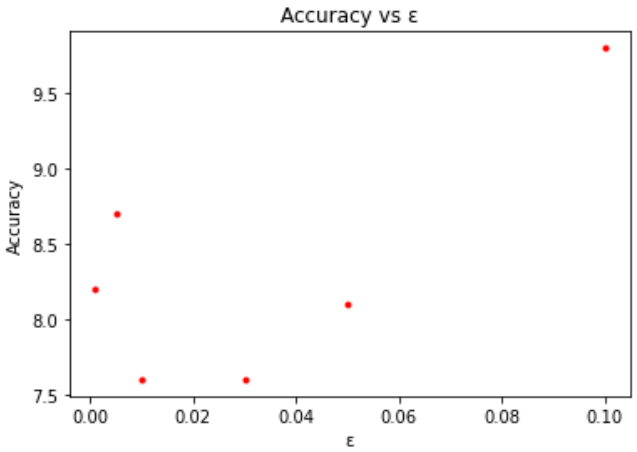
- **Task 1**



	0.0001	0.001	0.005	0.01	0.05	0.1	1.0
AE	10313.74	998.17	193.80	93.11	20.87	9.58	0.96
MSE	223955983.01	1888979.94	75621.18	16422.07	844.76	188.63	1.77

As the epsilon value increases our errors drop significantly in Laplace noise. This means that we get closer and closer to the actual value of the data. This is good for utility because we achieve a noisy data, but it is closer to the actual data.

- **Task 2**



Exponential Mechanism uses a probabilistic approach to the non-numeric resulting queries and because of this approach as shown in the table above its accuracy is bad. I expected its accuracy would be much higher than the resulting table. Maybe the problem is with picking the random output.

Part 3

	0.1	0.5	1.0	2.0	4.0	6.0
GRR	19716.35	19143.18	18016.82	14399.41	4787.88	801.29
RAPPOR	215025.47	193129.82	166481.12	118836.76	52580.53	20955.29
OUE	208830.06	162929.35	111870.12	42275.12	11333.47	13756.76

From the table we can see that GRR results in a much lower error rate than RAPPOR and OUE. But in terms of privacy GRR is bad because adversaries can determine, with very high accuracy (as seen on the table it results in lower error value with increased epsilon value), the resulting aggregate statistics of the users. RAPPOR and OUE implements a much better probabilistic model than GRR which results in higher error values for each epsilon value. This means that privacy of the users is preserved. With increased values of epsilon each algorithm results in a lower error value.