RWTH Aachen
Lehrstuhl für Informatik 9
Prof. Dr. T. Seidl

January 25th, 2016

Exercise 11 for the lecture
Data Mining Algorithms
WS 2015/2016

Hand in your solutions on February 1st before the lecture. The tutorial for this exercise will be held on February 5th.

Exercise 11.1) Privacy Preserving Data Mining — 8=2+2+4 points

a) Provide two objectives of Privacy Preserving Data Mining

b) Describe the relationship between Privacy and Data Utility (In 3 sentences)

c) Outline two major privacy paradigms and give two differences between them.

Exercise 11.2) *K*-Anonymity Privacy Definition — 8=4+4 points

Consider the table below which consists of sensitive medical record from the hospital.

Key Attribute	Quasi-Identifier			Sensitive Attribute
Name	**Gender**	**Age**	**Zip Code**	**Disease**
Alice	F	29	52066	Breast Cancer
Jane	F	27	52064	Breast Cancer
Jones	M	21	52076	Lung Cancer
Frank	M	35	52072	Heart Disease
Ben	M	33	52078	Fever
Betty	F	37	52080	Nose Pains

a) What is k-Anonymity and how does it ensures privacy

b) Make the table 2-Anonymous using only suppression

c) Make the table 2-Anonymous using generalization and suppression

d) Provide two shortcomings of k-Anonymity

e) What is l-Diversity and provide an example of a table that does not satisfy l-Diversity?

f) What is the difference between l-Diversity and t-Closeness?

Exercise 11.3) Differential Privacy **7 = 2+2+1+2 points**

a) Explain the core idea of differential privacy

b) Describe three kinds of noise that are employed to perturb data in differential privacy

c) Why is Laplace noise suitable for data perturbation?

d) Explain exponential mechanism (3 sentences).