Indian Institute of Technology Kharagpur
Intelligent Systems (CS60036)
Mid-Semester Examination, Spring 2017-18

Full Marks: 75 **Time: 2 hours**

All parts of the same question must be answered together
Be precise in your answers, and state any assumptions made
If there are multiple ways to perform a computation, state which one you are using

Question 1 [3+6+2+6+3 = 20 marks]

(a) With an example, define the following terms: (i) Favoritism, (ii) Direct Discrimination, (iii) Indirect Discrimination.

(b) List three possible sources of algorithmic bias, and explain how they can potentially lead to discrimination of certain societal groups.

(c) Explain the differences between disparate treatment and disparate impact.

(d) Using the contingency table below, define the following discrimination measures:
(i) Risk Difference, (ii) Odds Ratio and (iii) Extended Chance.

	benefit		
group	denied	granted	
protected	a	b	n_1
unprotected	c	d	n_2
	m_1	m_2	n

(e) How can consistency score measure individual discrimination? Can it be applied to measure group discrimination? If not, why?

Question 2 [7+3 =10 marks]

(a) Consider the following decision records of hiring decisions made in a company, based on the gender and qualification of the applicants.

Gender	Qualification	Hiring Decision
Male	Graduate	Yes
Female	Undergraduate	No
Female	Graduate	Yes
Male	Graduate	Yes
Male	Undergraduate	Yes
Female	Undergraduate	No

Female	Undergraduate	No
Female	Graduate	No
Male	Undergraduate	No
Male	Graduate	Yes
Female	Graduate	No

Check and justify whether the following association rules are 1.6-discriminatory:
(i) "Female, Graduate -> No"
(ii) "Female, Undergraduate -> No".
(b) How can association rules be used to consider genuine occupational requirement while discovering discrimination?

Question 3 [10 marks]
Derive a lower bound for indirect α-discrimination of PD classification rules given information available in PND rules (γ, δ) and information available from background rules (β1, β2).
More specifically, let **D**, **B** → **C** be a PND classification rule, and let:
γ = conf (**D**,**B** → **C**)
δ = conf (**B** → **C**) > 0
Let **A** be a PD itemset and let β1, β2 such that:
conf (**A**,**B** → **D**) ≥ β1
conf (**D**,**B** → **A**) ≥ β2 > 0

Given

$$f(x) = \frac{\beta_1}{\beta_2}(\beta_2 + x - 1)$$

$$elb(x, y) = \begin{cases} f(x)/y & \text{if } f(x) > 0 \\ 0 & \text{otherwise} \end{cases}$$

Prove that
(i) 1− f(1−γ) ≥ conf(A,B → C) ≥ f(γ)
(ii) For α ≥ 0 ,if PD classification rule A,B → C is α-discriminatory, then elb(γ,δ) ≥ α

Question 4 [5+2+3 = 10 marks]
(a) Using the decision records given as part of Question 2(a) and utilizing the concepts of situation testing, check whether women are t-discriminated where t = 0.2 Assume the neighborhood size to be 2. [Hint: you should consider corresponding educational qualifications.]
(b) In practical setting, how is the value of t determined?

(c) How can k-NN be used for discrimination discovery?

Question 5 [5+3+7 = 15 marks]

(a) How can situational testing and bayesian network be used for discrimination prevention?

(b) What is a redlining attribute? How do bayesian networks deal with redlining attributes while discrimination discovery?

(c) How do we build Suppes-Bayes network? Explain the steps to detect discrimination using causal networks.

Question 6 [3+7 = 10 marks]

(a) Devise the relation between predictability and disparate impact.

(b) How can a classifier be setup to certify lack of disparate impact in the decision making.