Here are some sample exam questions. ***The objective is to show you the style of questions, not to provide any hints for potential questions in the exam.*** The answer of most questions can be found in our slides.

1. Explain the difference between supervised and unsupervised learning.
2. What is the difference between classification and prediction?
3. Why is interpretability an important evaluation criterion for classification analysis? Provide some examples in your explanation.
4. Briefly describe the process of classification analysis. Why is it called supervised learning? What are the potential privacy and security risks caused by classification analysis?
5. What is the difference between macro-average and micro-average?
6. Why is tree pruning useful in decision tree induction?
7. Suppose you are asked to build a prediction model for an insurance company. The delivered model should be able to explain why a specific case is detected as fraud.

The following table consists of training data from an employee database. The data have been generalized. For example, "31 ... 35" for *age* represents the age range of 31 to 35. For a given row entry, *count* represents the number of data tuples having the values for *department*, *status*, *age*, and *salary* given in that row.

department	*status*	*age*	*salary*	*count*
sales	senior	31...35	46K...50K	30
sales	junior	26...30	26K...30K	40
sales	junior	31...35	31K...35K	40
systems	junior	21...25	46K...50K	20
systems	senior	31...35	66K...70K	5
systems	junior	26...30	46K...50K	3
systems	senior	41...45	66K...70K	3
marketing	senior	36...40	46K...50K	10
marketing	junior	31...35	41K...45K	4
secretary	senior	46...50	36K...40K	4
secretary	junior	26...30	26K...30K	6

Let *status* be the class label attribute.