

Q1-1: Which is a NOMINAL feature introduced in the lecture?

1. Cost $\in [0, 100]$
2. Awarded $\in \{\text{True}, \text{False}\}$
3. Steak $\in \{\text{Rare}, \text{Medium Rare}, \text{Medium}, \text{Medium Well}, \text{Well Done}\}$
4. Attitude $\in \{\text{strongly disagree}, \text{disagree}, \text{neutral}, \text{agree}, \text{strongly agree}\}$

Q1-1: Which is a NOMINAL feature introduced in the lecture?

1. Cost $\in [0, 100]$
2. Awarded $\in \{\text{True, False}\}$
3. Steak $\in \{\text{Rare, Medium Rare, Medium, Medium Well, Well Done}\}$
4. Attitude $\in \{\text{strongly disagree, disagree, neutral, agree, strongly agree}\}$



Q1-2: What is the dimension of the feature space?

The CIFAR-10 dataset contains 60,000 32x32 **color** images in 10 different classes.

(convert each data to a vector)

1. 10
2. 60,000
3. 3072
4. 1024

Q1-2: What is the dimension of the feature space?

The CIFAR-10 dataset contains 60,000 32x32 **color** images in 10 different classes.

(convert each data to a vector)

1. 10
2. 60,000
3. 3072
4. 1024



Every color image has 3 channels (RGB) and 32*32 pixels, so the dimension is $3*32*32=3072$.

Q2-1: Which generally is NOT a supervised learning task?

1. Binary classification
2. Email spam detection
3. Handwritten recognition
4. Eigenvalue calculation

Q2-1: Which generally is NOT a supervised learning task?

1. Binary classification
2. Email spam detection
3. Handwritten recognition
4. Eigenvalue calculation



Eigenvalue calculation is a mathematical problem, and we do not have any label for this problem.

Q2-2: Are these statements true or false?

(A) Instances from time series are independent and identically distributed.

(B) The primary objective of supervised learning is to find a model that achieves the highest accuracy on the training data.

1. True, True
2. True, False
3. False, True
4. False, False

Q2-2: Are these statements true or false?

(A) Instances from time series are independent and identically distributed.

(B) The primary objective of supervised learning is to find a model that achieves the highest accuracy on the training data.

1. True, True
2. True, False
3. False, True
4. False, False



- (A) Instances from time series usually have dependencies on the previous instances.
- (B) The primary objective of supervised learning is to find a model that generalizes.

Q3-1: Which generally is NOT an unsupervised learning task?

1. Principal component analysis
2. Fraud detection
3. CIFAR-10 image classification
4. Community detection

Q3-1: Which generally is NOT an unsupervised learning task?

1. Principal component analysis
2. Fraud detection
3. CIFAR-10 image classification
4. Community detection



1. Principal component analysis is a problem of dimensionality reduction.
2. You can think fraud detection as an anomaly detection problem.
3. CIFAR-10 image classification is a classification task for labeled image data.
4. Community detection is some clustering problem.

Q3-2: Are these statements true or false?

(A) We can use unsupervised learning methods to do image segmentation.

(B) Unsupervised learning methods can reveal structures of the data.

1. True, True
2. True, False
3. False, True
4. False, False

Q3-2: Are these statements true or false?

(A) We can use unsupervised learning methods to do image segmentation.

(B) Unsupervised learning methods can reveal structures of the data.

1. True, True



2. True, False

3. False, True

4. False, False

(A) There are lots of unsupervised learning methods to do image segmentation, such as graph cuts by energy minimization.

(B) Unsupervised learning can discover interesting regularities/structures/patterns that characterize instances.