### 1 Which of the following is assumed when classifying using the Naive Bayes method?

- **A**. The features have strong correlation with each other
- **B**. The features are all normalized before classification
- **C RIGHT**. The features are independent/conditionally independent of each other
- **D**. The data is low dimensional in nature

### 2 The primary role of Principal component analysis is

- **A**. Classification
- **B**. Regression
- **C**. Clustering
- **D RIGHT**. Dimensionality Reduction

### 3 Which direction does PCA use to perform dimensionality reduction?

- **A RIGHT**. Direction of maximum variance
- **B**. Direction of minimum variance
- **C**. Direction of maximum mean
- **D**. Direction of minimum mean

### 4 Which of the following is/are true about PCA?

- **A RIGHT**. The principal components are orthogonal to each other
- **B ~~RIGHT~~**. There can be only one principal component in PCA
- **C**. There are always lesser principal components than initial dimensions
- **D**. None of the above

---

Accepted Answers:

- The principal components are orthogonal to each other
- There are always lesser principal components than initial dimensions

---

### 5

The given probability distribution is $P(X|\mu) = \mu^x(1-\mu)^{(1-x)}$. Which of the following is True?

- A. $E[x] = \mu$
- B. Variance, $var(x) = \mu(1-\mu)$
- C. (Cross) Entropy: $-\mu ln\mu - (1-\mu)ln(1-\mu)$
- **D RIGHT**. All the above
  - Bernoulli distributiont. But here the range of x is not given in the question.

---

The outcome of a series of N tosses is given by the vector X which contains either 1 or 0 in each element. Assuming that the probability of a heads is μ and that each toss is independent, the probability P (X|μ) is given by

By above data answer the following questions 6 &7

---

### 6.

What would be the best name for the distribution

- **A RIGHT**.

$$P(X|\mu) = \prod_{i=1}^{N} \mu^{x_i} (1 - \mu)^{(1-x_i)}$$

- **B**.

$$P(X|\mu) = \prod_{i=1}^{N} \mu^{x_i}$$

- **C**.

$$P(X|\mu) = \prod_{i=1}^{N} (1 - \mu)^{(1-x_i)}$$

- **D**.

None of the above

### 7

If we are given the outcomes of a series of N tosses, the maximum likelihood estimation of the parameter $\mu$

- **A**.

$$\mu_{ML} = \frac{1}{N} \sum_{n=1}^{N} {x_n}^2$$

- **B**.

$$\mu_{ML} = \frac{1}{N} \sum_{n=1}^{N} n x_n$$

- **C**.

$$\mu_{ML} = \frac{1}{N} \sum_{n=1}^{N} n {x_n}^2$$

- **D RIGHT**.

$$\mu_{ML} = \frac{1}{N} \sum_{n=1}^{N} x_n$$

### 8 Which of the following is true for the Gaussian distribution?

- **A**. The mean of the distribution is always zero
- **B RIGHT**.In higher dimensions, covariance is represented by a matrix
- **C RIGHT**.The Gaussian is always symmetric about the mean
- **D RIGHT**.The mean of the Gaussian is the same as its mode

### 9 During linear regression, the maximum likelihood estimate of the parameters would be

- **A**.Greater than the least squares parameters
- **B**.Lesser than the least squares parameters
- **C RIGHT**.Same as the least squares parameters
  - In a linear model, if the errors belong to a normal distribution the least squares estimators are also the maximum likelihood estimators.
- **D**.Have no connection with the least squares parameters

### 10 Suppose you wanted to model the probability distribution of a set of people liking a particular dish. Which form of distribution would be most apt? (Hint: They can either like the dish or they can't.)

- **A**.Normal Distribution
- **B RIGHT**.Bernoulli Distribution
- **C**.Exponential Distribution
- **D**.Log-Normal Distribution