1. There are \( n \) binary features: \( \{ x_1, x_2, ..., x_n \} \). The hypothesis space consists of all disjunctions of size \( k \) (exactly) where hypotheses are of the form: \( z_{i_1} \lor z_{i_2} \lor ... \lor z_{i_k} \), where, \( i_1, i_2, ..., i_k \in \{ 1, 2, ..., n \} \) and \( z_{i_j} \in \{ x_{i_j}, \neg x_{i_j} \} \). What is the size of the hypothesis space?

Options:

- \( 2^k \cdot \binom{n}{k} \)
- \( k \cdot n \)
- \( \binom{n}{k} \)
- \( k \cdot \binom{n}{k} \)

Ans: \( 2^k \cdot \binom{n}{k} \)

2. The boolean function \( f \) over \( n \) variables: \( \{ x_1, x_2, ..., x_n \} \) is true if any \( m \) of the \( n \) variables are true. Which of the following is the correct linear representation for it?

Options:

- \( x_1 + x_2 + ... + x_n \geq m \)
- \( x_1 + x_2 + ... + x_m \geq n \)
- \( x_1 \cdot x_2 \cdot x_3 \cdot ... \cdot x_n \geq m \)
- \( x_1 \cdot x_2 \cdot x_3 \cdot ... \cdot x_m \geq n \)

Ans: \( x_1 + x_2 + ... + x_n \geq m \)

3. Which gradient technique is more suitable when the data is too large to bring into memory at once?

Options:

- Full Batch Gradient Descent
- Stochastic Gradient Descent

Ans: Stochastic Gradient Descent

4. You run gradient descent for \( 10 \) iterations with the value of learning rate \( \alpha = 0.5 \) and measure the loss \( E(w) \) after each iteration. You find the loss decreases quickly and then levels off (remains almost unchanged for the last 5 iterations). Which of the following seems plausible based on this?

Options:
• The current value of the learning rate is an effective choice and you should stop learning- you have converged
• It is promising to try a larger value of the learning rate
• It is promising to try a smaller value of the learning rate
• The current value of the learning rate is an effective choice and you should continue learning for a few more iterations

**Ans:** The current value of the learning rate is an effective choice and you should stop learning- you have converged

5. While learning a linear separator using a continuous and differentiable loss function $E(w)$, the batch gradient descent algorithm is guaranteed to find which of the following? Assume that the step size is set to an appropriate value.

**Options:**

- One of the local minima of $E(w)$
- The global minima of $E(w)$
- The global maxima of $E(w)$
- One of the local maxima of $E(w)$

**Ans:** One of the local minima of $E(w)$

6. Which of the following is NOT an example of binary classification?

**Options:**

- Predicting the part-of-speech of a given word in a sentence
- Predicting if an incoming email is spam or not
- Predicting whether a book review is positive or negative
- Predicting if an ad will get clicked or not

**Ans:** Predicting the part-of-speech of a given word in a sentence

7. Consider the problem of determining the part-of-speech tag of a single word in a given sentence. Which of the following is correct:

**Options:**

- Instance Space : Set of all words ; Label Space: Part-of-speech tags ; Hypothesis Space : Linear classifier
- Instance Space : Linear classifier ; Label Space: Set of all words ; Hypothesis Space : Part-of-speech tags
- Instance Space :Part-of-speech tags ; Label Space: Set of all words ; Hypothesis Space : Linear classifier
- Instance Space :Set of all words ; Label Space: Linear classifier ; Hypothesis Space : Part-of-speech tags
**Ans:** Instance Space: Set of all words; Label Space: Part-of-speech tags; Hypothesis Space: Linear classifier