1. Consider the following sample drawn from a Normal Distribution with unknown mean and unknown variance. 1 point

$$S = \{-1, 0, 1, 2\}$$

Select the Normal Distribution that is most likely to have been generated from sample S.

Hint: You may use the population variance instead of the sample variance in your calculations. Remember the notation for the Normal Distribution. $\mathrm{N}(\mu, \sigma^2)$*, where* μ *is the mean and* σ^2 *is the standard deviation.*

- ○ $\mathrm{N}(1.25, 0.5) = \mathrm{N}(1.12, 0.71^2)$
- ○ $\mathrm{N}(1, 0) = N(1, 0^2)$
- ◉ $\mathrm{N}(0.5, 1.25) = \mathrm{N}(0.5, 1.12^2)$
- ○ $\mathrm{N}(0, 1) = \mathrm{N}(0, 1^2)$

2. Which of the following statements about the Central Limit Theorem (CLT) is true? 1 point

- ○ The Central Limit Theorem states that the population mean will always be 0.
- ◉ The Central Limit Theorem states that the distribution of sample means approaches a normal distribution as the sample size increases, regardless of the shape of the population distribution.
- ○ The Central Limit Theorem suggests that the mean of a sample is always equal to the mean of the population.
- ○ The Central Limit Theorem only applies to populations that already follow a normal distribution. It has limited relevance in cases where the population distribution is skewed or has outliers.

3. Which of the following methods can be used to estimate a population's variance, mean, and proportion? 1 point

- ○ Sample mean
- ○ Sample variance
- ◉ Point estimation
- ○ Regression analysis

4. Suppose you flip a coin 10 times and obtain 6 heads and 4 tails. What function needs to be maximized to find the maximum likelihood estimate of the probability of getting heads on a single coin toss? Let p be the probability of getting heads. 1 point

- ○ $L(p) = p^{1/6}(1-p)^{1/4}$
- ◉ $L(p) = p^{6}(1-p)^{4}$
- ○ $L(p) = p^{4}(1-p)^{6}$
- ○ $L(p) = p^{10}(1-p)^{0}$

5. Suppose you have a dataset of points and want to fit a line to best represent the relationship between the variables. Which of the following statements about linear regression is true? 1 point

- ◉ Linear regression minimizes the sum of squared distances between the points and the fitted line, providing the best fit to the data.
- ○ Linear regression aims to maximize the sum of squared distances between the points and the fitted line.
- ○ Linear regression is unrelated to finding the best fit line for a given dataset.
- ○ Linear regression connects two random points in the dataset to fit the data.

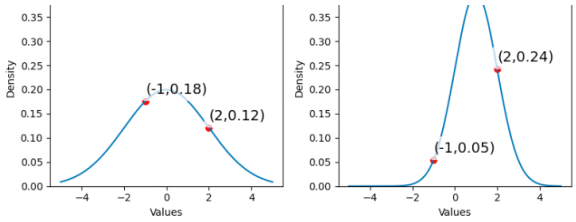
6. What is the purpose of regularization in machine learning? 1 point

- ○ Regulation favors more complex models to increase performance on the training data.
- ◉ Regularization prevents overfitting by penalizing models with large coefficients or weights.
- ○ Regularization is used to increase the training error of a model, which can improve its generalization performance.
- ○ Regularization is used to improve the interpretability of a model by reducing its complexity.

7. Assume you have a dataset that generates a model $M = 4x^4 + 3x^2 + 1$ that best fits the data. What is the L2 regularization error value for the model M? 1 point

- ○ 8
- ◉ 25
- ○ 26
- ○ 144

8. You have observations of the numbers $[-1, 2]$ and you want to determine which distribution they could have been sampled from. The first distribution is a normal distribution $N(0, 2^2)$ with a $\mu = 0$ and $\sigma = 2$. The second distribution is a normal distribution $N(1, 1^2)$ where $\mu = 1$ and $\sigma = 1$. 1 point



Which one is more likely to have generated the sample?

- (●) $N(0, 2^2)$
- ( ) $N(1, 1^2)$
- ( ) Cannot be determined.

9. Which of the following best describes the way "priors" are used in Bayesian statistics? 1 point

- ( ) After collecting data, prior beliefs are used to adjust the values of that data to better align with the patterns you expected to observe.
- ( ) Priors are used before data is available to assist in making conclusions. After collecting data, priors can be discarded.
- ( ) Priors are used to generate data in instances when direct observation of a phenomenon is impossible.
- (●) Prior beliefs are updated based on how well they align with the data observed.

10. Assume two Bayesian statisticians find a coin on the street and are trying to determine the likelihood that this coin will land on heads when flipped. 1 point

Bayesian 1 strongly believes that most coins are fair and begins wth priors that are heavily concentrated around $P(H) = 0.5$.

Bayesian 2 assumes they know nothing about coins and begins with uniform priors with equal likelihood of every $P(H)$ between 0 and 1.

In order to update their beliefs, they flip the coin 10 times, and get 3 heads and 7 tails.

Which of the following is the most likely value of their MAP beliefs once they've accounted for this data?

- ( ) Bayesian 1: $P(H) = 0.49$
  Bayesian 2: $P(H) = 0.30$
- (●) Bayesian 1: $P(H) = 0.51$
  Bayesian 2: $P(H) = 0.30$
- ( ) Bayesian 1: $P(H) = 0.30$
  Bayesian 2: $P(H) = 0.30$
- ( ) Bayesian 1: $P(H) = 0.30$
  Bayesian 2: $P(H) = 0.49$