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

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

Correct
Nice job! As you saw in the lecture MLE: Gaussian Example, the most probable Normal Distribution is the one with a mean equal to the mean of the data and a variance equal to the population variance of the data. In this case, the mean of S is 0.5 and its population variance is 1.25, leading to a standard deviation of 1.12, rounded to 2 decimal places.

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

- ○ 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.
- ○ The Central Limit Theorem suggests that the mean of a sample is always equal to the mean of the population.
- ◉ 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 states that the population mean will always be 0.

Correct
Nice job! The Central Limit Theorem (CLT) states that the distribution of sample means approaches a normal distribution as the sample size increases.

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

1 / 1 point

- [ ] Sample mean
- [ ] Sample variance
- [x] Point estimation
- [ ] Regression analysis

Correct
Point estimation is a statistical method that estimates population parameters such as mean, variance, and proportion based on a sample.

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 / 1 point

- [ ] $L(p) = p^{1/6}(1-p)^{1/4}$
- [x] $L(p) = p^6(1-p)^4$
- [ ] $L(p) = p^4(1-p)^6$
- [ ] $L(p) = p^{10}(1-p)^0$

Correct
That's right! We can use Bernoulli's distribution to find the likelihood function for this problem:
$L(p) = p^6 * (1-p)^4$.

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 / 1 point

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

Correct
Linear regression minimizes the sum of squared distances between the points and the fitted line, providing the best fit to the data.

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

1 / 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.

Correct
That's right! Regularization aims to balance fitting the data well and avoiding overly complex models.

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?

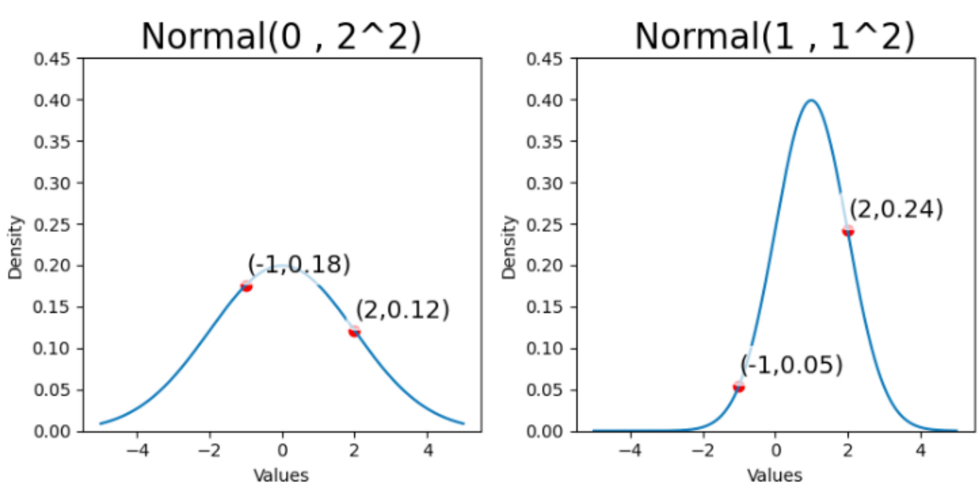
1 / 1 point

- ○ 8
- ◉ 25
- ○ 26
- ○ 144

Correct
Great job! The L2 regulation error is the sum of the squares of the coefficients, not including the last constant. Assuming the model can be represented by the polynomial $y = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x^1 + a_0$, the L2 regularization error is $a_n^2 + a_{n-1}^2 + \cdots + a_1^2$.

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 / 1 point



Which one is more likely to have generated the sample?

- [x] $N(0, 2^2)$
- [ ] $N(1, 1^2)$
- [ ] Cannot be determined.

Correct
$\text{Likelihood}(N(0, 2^2)) = 0.18 \cdot 0.12 = 0.216$ and
$\text{Likelihood}(N(1, 1^2)) = 0.05 \cdot 0.24 = 0.12$. Since
$\text{Likelihood}(N(0, 2^2)) > \text{Likelihood}(N(1, 1^2))$, The distribution $N(0, 2^2)$ is most likely to have generated the sample $[-1, 2]$. For further explanation, watch MLE: Gaussian Example.

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

1 / 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.
- [x] Prior beliefs are updated based on how well they align with the data observed.

Correct
As the term 'prior' implies, it reflects initial beliefs about the distribution of a parameter before observing any data. However, in Bayesian statistics, the strength of these prior beliefs is modified through a systematic update process driven by observed data. This updating is done through Bayes' theorem, ensuring that prior beliefs are refined and aligned with the information contained from the data. It's crucial to note that in Bayesian analysis, data is **never** discarded - it plays a central role in adjusting and shaping the prior distribution to better reflect the underlying reality.

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 / 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: $P(H) = 0.49$

Correct

The first Bayesian began with strongly held beliefs and so will only update them slightly in response to this data.

The second Bayesian makes the same conclusion as MLE, or in this case 3 out of 10, or 0.30.