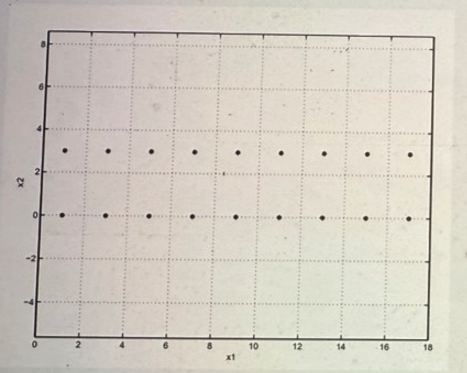
**1. Decision Trees I**

**Question:**  
Suppose you have a fixed dataset and want to train a decision tree for two-class classification. If you increase the maximum depth of the decision tree, which of the following are possible effects?

**Pick ONE OR MORE options:**

1. The test accuracy goes up.
2. The time to classify a test point increases.
3. The training accuracy goes down.
4. The number of pure leaves is reduced.

**2. K-means I**

****

**Question:**  
Consider the data plotted in the Figure above, which consists of two rows of equally spaced points. If k-means clustering (k = 2) is initialized with the two points whose coordinates are (9, 3) and (11, 3), what are the coordinates of the final clusters obtained (after the algorithm converges)?

**Pick ONE option:**

1. (9, 3) and (11, 0)
2. (6, 0) and (14, 3)
3. (6, 3) and (14, 0)
4. (9, 0) and (11, 3)

**3. Overlap Distributions**

**Question:**  
Suppose X is uniformly distributed over [0, 4] and Y is uniformly distributed over [0, 1]. Assume X and Y are independent. Then, P(max(X, Y) > 3) is equal to:

**Pick ONE option:**

1. 0
2. 1/4
3. 1/2
4. 3/4

**4. Random Forest I**

**Question:**  
Why would we use a random forest instead of a decision tree?

**Pick ONE OR MORE options:**

1. For lower training error.
2. To better approximate posterior probabilities.
3. To reduce the variance of the model.
4. For a model that is easier for a human to interpret.

**5. Convexity I**

**Question:**  
Assume you have a function *f*, which can be made up of many smaller, simpler convex functions after some operations. Which of the following operations will preserve convexity? Check all that apply.

**Pick ONE OR MORE options:**

1. lasso
2. minimization
3. nonnegative weighted sum
4. perspective
5. composition
6. pointwise maximum and supremum
7. maximum likelihood
8. composition with affine function
9. maximum entropy

**6. Decision Trees II**

**Question:**  
Which of the following would be a reasonable cost function for choosing splits in a decision tree for two-class classification, where *p* is the fraction of points belonging to one class in a specified treenode?

**Pick ONE OR MORE options:**

1. 0.5 − |p − 0.5|
2. −p log₂p − (1 − p) log₂(1 − p)
3. p log₂p + (1 − p) log₂(1 − p)
4. p(1 − p)

**7. Hypothesis Testing I**

**Question:**  
Two types of errors associated with hypothesis testing are Type I and Type II. Type II error is committed when:

**Pick ONE option:**

1. We reject the null hypothesis whilst the alternative hypothesis is true
2. We reject a null hypothesis when it is true
3. We accept a null hypothesis when it is not true

**8. Neural Network I**

**Question:**  
Which of the following are true in general for backpropagation?

**Pick ONE OR MORE options:**

1. It is a dynamic programming algorithm
2. The weights are initially set to zero
3. Its running time grows exponentially in the number of layers
4. Some of the derivatives cannot be fully computed until the backward pass

**9. Three Points or Two Points**

**Question:**  
You are the best player in your team, you have the ball. Your team is down two points and the clock has three seconds left. Historically you have a 40% probability of getting the 3 point shot and a 70% probability of getting the 2 point shot. If you take the two pointer, it produces a tie and puts you in overtime. And in overtime, we assume your team has a fifty-fifty chance of winning.  
Which is better? Taking the 3 pointer to get the sure win or taking the two pointer?

**Pick ONE option:**

1. Take the 3 point shot
2. Take the 2 point shot

**10. Rank II**

**Question:**  
A low-rank approximation of a matrix can be useful for:

**Pick ONE OR MORE options:**

1. Removing noise
2. Discovering latent categories in the data
3. Filling in unknown values
4. Matrix compression

**11. Poisson**

**Question:**  
Earthquakes in Sumatra occur according to a Poisson process of rate λ = 2/year. Conditioned on the event that exactly two earthquakes take place in a year, what is the probability that both earthquakes occur in the first three months of the year?  
(For simplicity, assume all months have 30 days, and each year has 12 months, i.e., 360 days).

**Pick ONE option:**

1. 1/2
2. 1/4
3. 1/16
4. 1/9

**12. K-Nearest Neighbors II**

**Question:**  
In which of the following cases should you prefer k-nearest neighbors over k-means clustering? For all the four options, you have access to images X₁, X₂, ..., Xₙ ∈ ℝᵈ.

**Pick ONE OR MORE options:**

1. You do not have access to labels. You want to find out if any of the images are very different from the rest, i.e., are outliers.
2. You have access to labels y₁, y₂, ..., yₙ telling us whether image i is a cat or a dog. You want to find out whether the distribution of cats is unimodal or bimodal. You already know that the distribution of cats has either one or two modes, but that's all you know about the distribution.
3. You have access to labels y₁, y₂, ..., yₙ telling us whether image i is a cat or a dog. You want to find out whether a new image z is a cat or a dog.
4. You have access to labels y₁, y₂, ..., yₙ telling us whether image i is a cat or a dog. Given a new image z, you want to approximate the posterior probability of z being a cat and the posterior probability of z being a dog.

**13. Regularization II**

**Question:**  
Consider the minimizer w\* of the l₂-regularized least squares objective  
J(w) = ‖Xw − y‖² + λ‖w‖² with λ > 0.  
Which of the following are true?

**Pick ONE option:**

1. Xw² = y
2. w\* = X⁺y, where X⁺ is the pseudoinverse of X
3. w\* exists if and only if XᵀX is nonsingular
4. The minimizer w\* is unique

**14. White Noise I**

**Question:**  
A white noise process will have:  
(i) A zero mean  
(ii) A constant variance  
(iii) Autocovariances that are constant  
(iv) Autocovariances that are zero except at lag zero

**Pick ONE option:**

1. (ii) and (iv) only
2. (i) and (iii) only
3. (i), (ii), and (iii) only
4. (i), (ii), (iii), and (iv)

**15. Eigenvalues II**

**Question:**  
Given any matrix X, (XᵀX + λI)⁻¹ for λ ≠ 0 always exists.

**Pick ONE option:**

1. True
2. False

**16. Statistical Bias**

**Question:**  
"When the Brodie helmet was introduced during WWI, there was a dramatic rise in field hospital admissions of severe head injury victims. This led army command to consider redrawing the design until a statistician remarked that soldiers who might previously have been killed by certain shrapnel hits to the head (and therefore never showed up in a field hospital), were now surviving the same hits, and thus made it to a field hospital."

Which type of statistical bias is evident in this example?

**Pick ONE option:**

1. Sampling Bias
2. Confirmation Bias
3. Survivorship Bias
4. Susceptibility Bias

**17. Tails I**

**Question:**  
The choice of one-tailed test and two-tailed test depends upon:

**Pick ONE option:**

1. None of these below
2. Alternative hypothesis
3. Null hypothesis
4. Composite hypotheses

**18. K-Nearest Neighbors I**

**Question:**  
What tends to be true about increasing the k in k-nearest neighbors?

**Pick ONE OR MORE options:**

1. The decision boundary tends to get smoother
2. The bias tends to increase
3. The variance tends to increase
4. As the number of sample points approaches infinity (with n/k → ∞), the error rate approaches less than twice the Bayes risk (assuming training and test points are drawn independently from the same distribution)

**19. Correlation & Covariance**

**Question:**  
Choose the correct statement:

**Pick ONE option:**

1. Correlation and covariance range between -1 and +1
2. Correlation ranges between -1 and +1. Covariance ranges between -infinity and +infinity
3. Correlation and covariance range -infinity and +infinity
4. Correlation ranges between -infinity and +infinity. Covariance ranges between -1 and +1

**20. Plant Operations**

**Question:**  
You are given a budget of $150,000 to operate a cement plant. You have ten customers coming to you asking you to make cement for them for their projects. Each project has forecast the potential revenue you can get and the potential cost. Below is the breakdown:

| **Project** | **Revenue ($)** | **Cost ($)** |
| --- | --- | --- |
| P1 | 30,000 | 5,000 |
| P2 | 20,000 | 2,500 |
| P3 | 250,000 | 100,000 |
| P4 | 250,000 | 75,000 |
| P5 | 700,000 | 65,000 |
| P6 | 600,000 | 12,000 |

You are asked to maximize the ROI of each project. Given your budget constraint, which projects will you choose?

**Pick ONE option:**

1. P1, P2, P3
2. P2, P4, P5
3. P3, P5, P6
4. All of them

**21. LLN**

A white text with black text

AI-generated content may be incorrect.

**Pick ONE option:**

1. Statement One: False Statement Two: False Statement Three: False
2. Statement One: False Statement Two: False Statement Three: True
3. Statement One: False Statement Two: True Statement Three: True
4. Statement One: True Statement Two: True Statement Three: False
5. Statement One: True Statement Two: False Statement Three: True
6. Statement One: True Statement Two: True Statement Three: True

**22. Hessian**

**Question:**  
Determine the definiteness of the Hessian Matrix of the following function:  
f(x,y,z) = ln(x) + ln(y) + ln(z)

**Pick ONE option:**

1. Indefinite
2. Positive definite
3. Negative definite
4. None of the above

**23. Vector and Scalar Fields**

Which of the following are true?

**Pick ONE OR MORE options:**

1. Gradient: Vector of the first order derivatives of a scalar field
2. Hessian: Vector of the first order derivatives of a scalar field
3. The Jacobian is the Hessian of the gradient of a function that maps from N-dimensions to 1-Dimension
4. Jacobian: Matrix of the second order mixed partials of a scalar field
5. Hessian: Matrix of the second order mixed partials of a scalar field
6. Jacobian: Matrix of the gradients for components of a vector field
7. The Hessian is the Jacobian of the gradient of a function that maps from N-dimensions to 1-Dimension

**24. Drug Trial**

**Question:**  
A pharmaceutical company is investigating a new drug (Drug A) as a potential replacement for an old drug (Drug B). They make several statements. Which of these statements most relates to whether we can determine the relationship between the drug and survival rate is a cause-and-effect relationship?

**Pick ONE OR MORE options:**

1. Patients using Drug A have a survival rate that is significantly higher than patients using Drug B
2. The patients in the study were aged 20–40 with early-stage disease progression and were randomly assigned by the experimenter to take one of two drugs
3. There were 100 patients in each group
4. None of the above

**25. Lagrange Multipliers**

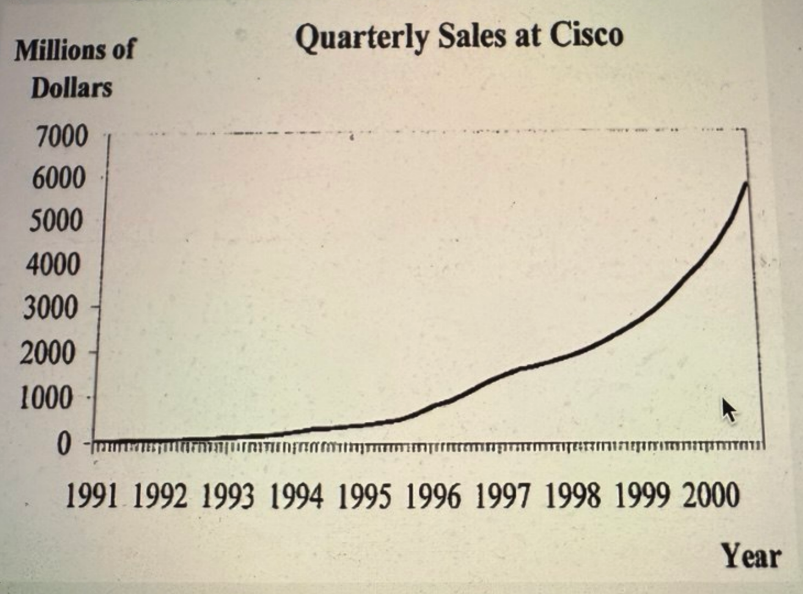
**Question:**  
Find the x and y value that maximizes the function z = 4 - 5x + 6y subject to the constraint x + y = 30

**Pick ONE option:**

1. x = 15, y = 15
2. x = 17, y = 13
3. x = 20, y = 10
4. No maxima exists

**26. AR models I**

**The following figure shows the quarterly sales of Cisco Systems from Q1 of 1991 to Q4 of 2000.**



The following table gives the regression statistics from estimating the model:

**Change in the Natural Log of Sales for Cisco Systems**  
**Quarterly Observations, Q3:1991 – Q4:2000**

**Regression Statistics**

* R-squared: 0.2899
* Standard error: 0.0408
* Observations: 38
* Durbin–Watson: 1.5707

**Regression Coefficients**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Coefficient** | **Standard Error** | **t-Statistic** |
| Intercept | 0.0661 | 0.0175 | 3.7840 |
| Δ ln(Sales t) | 0.4698 | 0.1225 | 3.8339 |

**Question:**  
Assuming the model is correctly specified, does this time series tend towards its mean reverting level?

**Pick ONE OR MORE options:**

1. Yes. This means the time series is covariance stationary, hence it tends towards its mean reverting level.
2. No. Even though the model is correctly specified, it does not mean the time series is covariance stationary. As a result mean reversion cannot be accurately assessed.
3. Yes. This means the time series has little correlation with its past, hence it tends towards its mean reverting level.
4. Yes. This means the time series at the log scale will reach an asymptote, hence it tends towards its mean reverting level.

**27. Pseudo Inverse**

**Question:**  
Suppose you are given a linear equation Ax = b where A is an (m x n) matrix and m > n. What are the possible next step(s) towards computing a pseudoinverse?

**Pick ONE OR MORE options:**

1. A⁻¹Ax = A⁻¹b
2. AᵀAx = Aᵀb
3. (AᵀA)⁻¹(AᵀA)x = (AᵀA)⁻¹Aᵀb
4. AxA⁻¹ = bA⁻¹

**28. K-means II**

**Question:**  
What is true of k-means clustering?

**Pick ONE OR MORE options:**

1. k-means is a supervised learning algorithm.
2. k-means clustering always converges to the same solution regardless of how clusters are initialized.
3. Increasing k can never increase the optimal value of the k-means cost function.
4. The k-medoids algorithm with the L1 distance is less sensitive to outliers than standard k-means with the Euclidean distance.

**29. Newton-Raphson II**

**Question:**  
Given function f(x) = |x² - 3| - 1 defined on ℝ: Which of the following is true?

**Pick ONE option:**

1. Newton-Raphson Method on minimizing gradients will always converge to the global optimum in one iteration from any starting location
2. Stochastic Gradient Descent will always converge to the global optimum in one iteration
3. The problem is nonconvex, so it is not feasible to find a solution.
4. All of the above
5. None of the above

**30. Seats in a plane**

**Question:**  
A line of 100 passengers is waiting to board a plane. They each hold a ticket to one of the 100 seats on that flight. The first person in line is crazy and will pick a random seat to occupy. All other passengers will go to their assigned seat unless it is occupied, then they choose a random unoccupied one. What is the probability that the last (100th) person to board the plane will sit in their proper seat (#100)?

**Pick ONE option:**

1. 0.1
2. 0.5
3. 0.99
4. 0.01

**31. Marbles**

**Question:**  
There are two boxes: The first one contains 900 red marbles and 100 blue marbles, the other one contains 500 red marbles and 500 blue marbles. If we pick one of the boxes randomly, and pick a marble, what is the probability that it is blue?

**Pick ONE option:**

1. 0.25
2. 0.3
3. 0.35
4. 0.4

**32. Feature Normalization**

**Question:**  
For which of the following does normalizing your input features influence the predictions?

**Pick ONE OR MORE options:**

1. Decision tree (with usual splitting method)
2. Neural Network
3. Lasso
4. Soft-margin SVM
5. All of the above

**33. Newton-Raphson I**

**Question:**  
Which of the following is true?

**Pick ONE OR MORE options:**

1. Newton’s Method typically is more expensive to calculate than gradient descent, per iteration
2. For quadratic equations, Newton’s Method typically requires fewer iterations than gradient descent
3. Gradient descent can be viewed as iteratively reweighted least squares
4. None of the above

**34. Sets II**

**Question:**  
Suppose you have two limit orders outstanding on two different stocks. The probability that the first limit order executes before the close of trading is 0.45.  
The probability that the second limit order executes before the close of trading is 0.20.  
The probability that the two orders both execute before the close of trading is 0.1.  
What is the probability that at least one of the two limit orders will execute before the close of trading?

**Pick ONE option:**

1. 0.75
2. 0.65
3. 0.55
4. 0.45

**35. Ball arrangements**

**Question:**  
There are 5 balls, 3 red, and 2 black. What is the total number of ways that a random ordering of the 5 balls does not have the 2 black balls next to each other?

**Pick ONE OR MORE options:**

1. 10
2. 8
3. 6
4. None of the above

**36. PCA II**

**Question:**  
Let X be a matrix with singular value decomposition X=UΣVᵀ. Which of the following are true for all X?

**Pick ONE OR MORE options:**

1. rank(X) = rank(Σ)
2. If all the singular values are unique, then the SVD is unique
3. The first column of V is an eigenvector of XᵀX
4. The columns of V (right-singular vectors) are eigenvectors of XᵀX
5. The columns of U (left-singular vectors) are eigenvectors of XXᵀX

**37. ARCH I**

**Question:**  
The following table shows the regression output for testing for ARCH(1) in the residuals from an AR(1) regression on first differences in the civilian unemployment rate (UER):  
Δ = Δ +

**Testing for ARCH(1)**  
**Squared Residuals from the Monthly Changes in Civilian Unemployment Rate**

**Regression Statistics**

* R-squared: 0.0007
* Standard error: 0.0208
* Observations: 57
* Durbin–Watson: 1.9783

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Coefficient** | **Standard Error** | **t-Statistic** |
| Intercept | 0.0146 | 0.0034 | 4.3468 |
|  | -0.0265 | 0.1348 | -0.1670 |

Using information in the table, is this regression free from ARCH?

**Pick ONE option:**

1. Yes, the t-statistic for εt−1​ is insignificant. Hence the regression is free from ARCH
2. No, the t-statistic for the intercept is highly significant. Hence the regression is impacted by ARCH

**38. Hypothesis Testing III**

**Question:**  
Suppose that you want to test the claim made by a startup that their transport network is able to complete a food delivery in under an hour. For a random sample of 24 food deliveries, the startup delivery completion time averages 63.2 minutes with a standard deviation of 7.7 minutes. (Assume a normal distribution).

What is the test statistic for this test?

**Pick ONE option:**

1. Z = 2.04
2. t = 1.79
3. t = 2.04
4. Z = 1.79

**39. Decision Trees III**

**Question:**  
We want to use a decision tree to classify the training points depicted.  
Which of the following decision tree classifiers is capable of giving 100% accuracy on the training data with four splits or fewer?

A graph with blue and orange dots

AI-generated content may be incorrect.

**Pick ONE OR MORE options:**

1. A standard decision tree with axis-aligned splits
2. Using PCA to reduce the training data to one dimension, then applying a standard decision tree
3. A decision tree with multivariate linear splits
4. Appending a new feature ∣x1∣ + ∣x2∣ to each sample point x, then applying a standard decision tree

**40. Hypothesis Testing II**

**Question:**  
In a two-tailed test when a Null Hypothesis is rejected for a True Alternative Hypothesis then it has

**Pick ONE option:**

1. Type 1 Error
2. Type 2 Error
3. No Error
4. Many errors

**41. Linear Models II**

**Question:**  
You have been assigned to analyze the automobile industry and you decide to model the monthly sales of lightweight vehicles to determine sales growth in this segment of the market. The following figure gives lightweight vehicle monthly sales (annualized) from January 1992 to December 2000.

A graph of a sales report

AI-generated content may be incorrect.

Monthly sales in the lightweight vehicle sector, Sales, have been increasing over time, but you suspect the growth rate of monthly sales is constant. Write the simplest time model, Salest, that is consistent with your belief.

**Pick ONE option:**

1. ln(Salesₜ) = b₀ + b₁t + εₜ
2. Salesₜ = b₀ + b₁t + εₜ
3. Salesₜ = b₀ + b₁t + b₂t² + εₜ
4. Salesₜ = b₀ + b₁t

**42. Statistical Significance**

**Question:**  
A researcher collected data that fit the criteria for a two-sided Z-test. He set the significance level at 0.05. He ran 80 trials and got a z-value of 1.7. This gave a p-value of 0.0892, so he could not reject the null hypothesis. Convinced that his alternative hypothesis was correct he ran 80 more trials. The combined data from the 160 trials now had a z-value of 2.1. He wrote a paper carefully describing his experiments and submitted it to the journal in part (a). Will the journal publish his results?

**Pick ONE option:**

1. Yes
2. No

**43. HMM**

**Question:**

Assume that the following sequences are very long and the pattern with spaces are repeated:  
Sequence 1: 00100100100 … 100  
Sequence 2: 1100100100 … 100

If we model each sequence with a different first-order HMM, what is the number of hidden states that a reasonable model selection method would report?

**Pick ONE option:**

1. Sequence 1 : 2, Sequence 2 : 2
2. Sequence 1 : 2, Sequence 2 : 3
3. Sequence 1 : 3, Sequence 2 : 4
4. Sequence 1 : 5, Sequence 2 : 6

**44. Permutations**

**Question:**  
Does a monkey have a better chance of rearranging "ACCLLUUS" to spell "CALCULUS" than of rearranging "AABEGLR" to spell "ALGEBRA"?

**Pick ONE option:**

1. True
2. False

**45. KKT I**

**Question:**  
Given the general problem:

  min f(x)  
  subject to   (x) ≤ 0, i = 1, ..., m  
         (x) = 0, j = 1, ..., r

The Karush-Kuhn-Tucker conditions or KKT conditions are:

A.  0 ∈ δf(x) + δ(x) + ​δ(x)   
B.   · (x) = 0 for all i  
C.  (x) ≤ 0, (x) = 0 for all i, j  
D.   > 0 for all i

What are the names of conditions A, B, C and D?

**Pick ONE OR MORE options:**

1. A. stationarity B. complementary slackness C. primal feasibility D. dual feasibility
2. A. complementary slackness B. primal feasibility C. dual feasibility D. stationarity
3. A. dual feasibility B. primal feasibility C. complementary slackness D. stationarity
4. A. primal feasibility B. dual feasibility C. complementary slackness D. stationarity

**46. Conditional Probability**

**Question:**  
If P(A|B) > P(A) and P(A|C) > P(A), then P(A|B,C) > P(A). True or False?

**Pick ONE option:**

1. True
2. False

**47. MLE I**

**Question:**  
Maximum likelihood estimation gives us not only a point estimate, but a distribution over the parameters that we are estimating. True or False?

**Pick ONE option:**

1. True
2. False

**48. Backprop I**

**Question:**  
Backpropagation is motivated by utilizing Chain Rule and Dynamic Programming to conserve mathematical calculations. True or False?

**Pick ONE option:**

1. True
2. False

**49. Documents I**

**Question:**  
Suppose we are trying to solve an active learning problem, where the possible inputs you can select form a discrete set. Specifically, we have a set of N unlabeled documents, Φ₁, Φ₂, ..., Φₙ, where each document is represented as a binary feature vector.

φᵢ = [φᵢ₁, φᵢ₂, ..., φᵢᵐ]ᵀ  
and φᵢⱼ = 1 if word j appears in the document and zero otherwise. Our goal is to quickly label these N documents with 0/1 labels. We can request a label for any of the N documents, preferably as few as possible. We also have a small number n of these documents already labeled to get us started. We use a logistic regression model to solve the classification task:

P(y = 1|Φ, w) = g(wᵀx)  
where g(·) is the logistic function. Note that we do not include the bias term.

Any word that appears in all the N documents would effectively provide a bias term for the logistic regression model. True or False?

**Pick ONE option:**

1. True
2. False

**50. Linear Matrix**

**Question:**  
It is known that matrix A ∈ ℝ³ˣ³ (a matrix consisting of real numbers with a 3×3 dimension) has eigenvalues 0, 1 and 2.  
Does the linear system Ax = b where (x; b ∈ ℝ³) have a solution?

**Pick ONE option:**

1. Yes. Because matrix A is not diagonalizable and b is in the column space of A
2. Yes. Because there are three distinct eigenvalues and b is in the column space of A
3. No. Because matrix A is not diagonalizable and b is not in the column space of A
4. No. Because there are three distinct eigenvalues and b is not in the column space of A

**51. Random Forest vs Adaboost**

**Question:**  
Which of the following are advantages to using AdaBoost with short trees (say, depth 4) over random forests with an equal number of tall trees (refined until the leaves are pure)?

**Pick ONE OR MORE options:**

1. AdaBoost is more robust against overfitting outliers in the training data.
2. AdaBoost is better at reducing variance than a random forest.
3. AdaBoost is better at reducing bias than a random forest.
4. AdaBoost is faster to train.

**52. Coins**

**Question:**  
You have 17 coins and I have 16 coins, we flip all coins at the same time. If you have more heads then you win, if we have the same number of heads or if you have less then I win. What's your probability of winning?

**Pick ONE option:**

1. 0.75
2. 0.25
3. 0.5
4. 0.125

**53. Cards**

**Question:**  
A standard deck of 52 cards contains 4 aces. Suppose we choose a random ordering (all 52! permutations being equally likely). Compute the probability that all the top four cards are aces.

**Pick ONE option:**

1. 1/13
2. 4!4!44!/52!
3. 4!48!/52!
4. 4/13

**54. Rank**

**Question:**  
Find the rank of the following matrix:

[1 3 21]

[5 7 49]

[2 5 35]

**Pick ONE option:**

1. 3
2. 2
3. 1
4. 0

**55. Confidence Intervals**

**Question:**  
A interval estimation is just the probability that your estimator (say 𝜃̂) is contained in a random interval with probability 1 - α. True or False?

**Pick ONE option:**

1. True
2. False

**56. Stationarity I**

**Question:**  
Which of the following conditions are necessary for a series to be classifiable as a weakly stationary process?  
(i) It must have a constant mean  
(ii) It must have a constant variance  
(iii) It must have constant autocovariances for given lags  
(iv) It must have a constant probability distribution

**Pick ONE option:**

1. (ii) and (iv) only
2. (i) and (iii) only
3. (i), (ii) and (iii) only
4. (i), (ii), (iii) and (iv)

**57. Inverse of Matrix**

**Question:**  
Given a symmetric matrix such that its inverse exists, what is the inverse of this matrix?

**Pick ONE option:**

1. A diagonal matrix
2. A symmetric matrix
3. An identity matrix
4. Cannot be determined

**58. Bivariates**

**Question:**  
In general, if two variables, X, Y are uncorrelated such that [ρ(X, Y) = 0], then they are independent. For example, the Bivariate Normal Distribution: when both random variables are uncorrelated they are also independent. True or False?

**Pick ONE option:**

1. True
2. False

**59. Neural Network III**

**Question:**  
Below are some choices you might make while training a neural network. Select all of the options that will generally make it more difficult for your network to achieve high accuracy on the test data.

**Pick ONE OR MORE options:**

1. Initializing the weights to all zeros
2. Normalizing the training data but leaving the test data unchanged
3. Using momentum
4. Reshuffling the training data at the beginning of each epoch

**60. SVD I**

**Question:**  
With the SVD, we write X = UDVᵀ. For which of the following matrices are the eigenvectors the columns of U?

**Pick ONE OR MORE options:**

1. XᵀX
2. XᵀX XᵀX
3. XXᵀ
4. XXᵀX

**61. Convexity II**

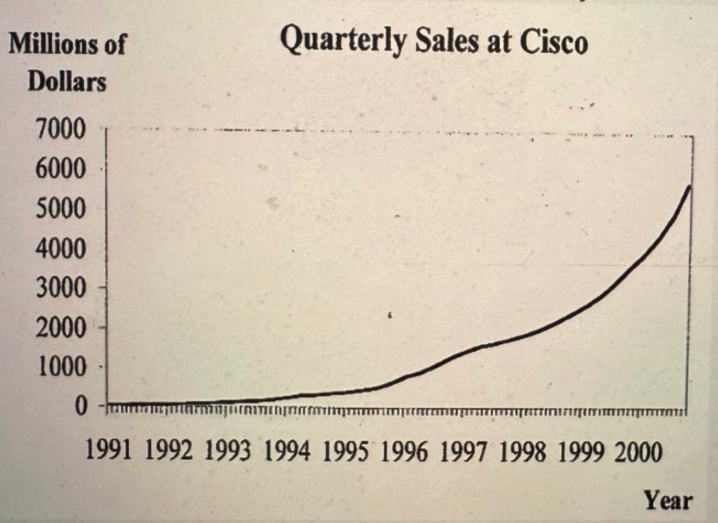
**Question:**  
A continuous, twice-differentiable function is convex if its Hessian

**Pick ONE OR MORE options:**

1. has only non-negative eigenvalues
2. is positive semi-definite
3. is negative semi-definite

**62. AR models II**

The following figure shows the quarterly sales of Cisco Systems from Q1 of 1991 to Q4 of 2000.



The following table gives the regression statistics from estimating the model:  
ln(Salesₜ) = b₀ + b₁ ln(Salesₜ₋₁) + εₜ

**Change in the Natural Log of Sales for Cisco Systems**  
**Quarterly Observations, Q3:1991 – Q4:2000**

**Regression Statistics**

* R-squared: 0.2899
* Standard error: 0.0408
* Observations: 38
* Durbin–Watson: 1.5707

**Regression Coefficients**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Coefficient** | **Standard Error** | **t-Statistic** |
| Intercept | 0.0661 | 0.0175 | 3.7840 |
| Δ ln() | 0.4698 | 0.1225 | 3.8339 |

**Question:**  
Which of the following describes the correct steps to determine if the AR(1) model is correct? Check the ones that apply.

**Pick ONE OR MORE options:**

1. First, we should determine whether the residuals from the AR(1) specification are serially uncorrelated. If the residuals are serially correlated, then we should try an AR(2) specification and then test the residuals from the AR(2) model for serial correlation. We should continue in this fashion until the residuals are serially uncorrelated.
2. We should determine whether the residuals from the AR(1) specification are serially uncorrelated. If the residuals are serially correlated, and since we are only assessing the AR(1) model, we can stop and conclude the model is not correct.
3. We must test for ARCH regardless whether seasonality is present.
4. Look for seasonality in the residuals. If seasonality is present, we should add a seasonal lag.
5. If no seasonality is present, we should test for ARCH. If ARCH is not present, we can conclude that the model is correctly specified.

**63. Independence**

**Question:**  
P(A) = P(A|B)P(B) means that A and B are independent. True or False?

**Pick ONE option**

1. True
2. False

**64. Neural Network II**

**Question:**  
Suppose our input is two-dimensional sample points, with ten non-exclusive classes those points may belong to (i.e., a point can belong to more than one class). To train a classifier, we build a fully-connected neural network (with bias terms) that has a single hidden layer of twenty units and an output layer of ten units (one for each class). Which statements apply?

**Pick ONE OR MORE options**

1. For the output units, softmax activations are more appropriate than sigmoid activations
2. For the hidden units, ReLU activations are more appropriate than linear activations
3. This network will have 270 trainable parameters
4. This network will have 240 trainable parameters

**65. Jobs I**

**Question:**  
Mary sends her resume to 1000 companies she finds on jobs.com. Each company responds with probability 3/1000 (independently of what all the other companies do). Let R be the number of companies that respond.  
Compute E(R), i.e. the expected value of R.

**Pick ONE option**

1. 3
2. 0.3
3. 0.03
4. 0.003

**66. Eigenvalues I**

**Question:**  
Given any matrix X, its singular values are the eigenvalues of XXᵀ and XᵀX.  
True or False?

**Pick ONE option**

1. True
2. False

**67. Dogs**

**Question:**  
Two dog owners are interested in whether their (shared) dog tends to run to a specific one of them first when they come home together. They walk through the door at the same time and record which person the dog runs to first. They repeat this procedure many times. What test should they use to ask if the dog runs to one of them more often than we would expect if the dog was choosing randomly between the owners?

**Pick ONE option**

1. Binomial test for a single proportion
2. One-sample t-test for a single mean
3. Z-test for a difference in two proportions
4. Two sample t-test for a difference in means

**68. Sets I**

**Question:**  
P(A) = P(A|B)P(B) means A and B are independent. True or False?

**Pick ONE option**

1. True
2. False

**69. Step size**

**Question:**  
You want to minimize a convex loss function f(x) using stochastic gradient descent. Given a random starting point, mark the condition that would guarantee that stochastic gradient descent will converge to the global optimum. Let ηₜ = step size at iteration t.

**Pick ONE OR MORE options**

1. ηₜ < 0
2. Constant step size ηₜ
3. Decreasing step size ηₜ = 1/√t
4. Decreasing step size ηₜ = 1/t²

**70. Estimators**

**Question:**  
If an estimator is consistent, it is also unbiased. Thus, maximum likelihood estimators are unbiased. True or False?

**Pick ONE option:**

1. True
2. False

**71. Duality**

**Question:**  
Which of the following statements about duality and SVMs is (are) true?

**Pick ONE OR MORE options:**

1. Complementary slackness implies that every training point that is misclassified by a soft-margin SVM is a support vector.
2. When we solve the SVM with the dual problem, we need only the dot product of xᵢ, xⱼ for all i, j, and no other information about the xᵢ.
3. We use Lagrange multipliers in an optimization problem with inequality (≤) constraints.
4. None of the above

**72. Elephant vs Donkey**

**Question:**  
A polling company finds that 60% of Democrats give a “favorable” rating to a senator and 52% of Republicans give a “favorable” rating for the same senator. The population size is rather large, around 200000. What should the company use if they want to know if this is a “statistically significant” finding? Pick the answer(s) that are appropriate.

**Pick ONE OR MORE options:**

1. Binomial test for a single proportion
2. One sample t-test for a single mean
3. Z-test for a difference in two proportions
4. Two sample t-test for a difference in means

**73. Dots and Squares**

**Question:**  
I have a square, and place three dots along the 4 edges at random. What is the probability that the dots lie on distinct edges?

**Pick ONE option:**

1. 3/8
2. 1/8
3. 3/4
4. 1/4

**74. Covid**

**Question:**  
Covid test kits are widely available these days, but they are not always accurate. Suppose 60% of the whole population have Covid and a diagnostic test has 99% accuracy. If a patient tests positive, what is the probability that they actually have the disease?

**Pick ONE option:**

1. 0.988
2. 0.902
3. 0.993
4. 0.989

**75. Linear Models I**

**Question:**  
The following figure compares the predicted civilian unemployment rate (PRED) vs the actual unemployment rate (UER) from January 1996 to December 2000. The predicted results come from estimating a linear time model:

​

A graph of unemployment rate

AI-generated content may be incorrect.

How appropriate is this model for predicting unemployment rate?

**Pick ONE OR MORE options:**

1. Very Appropriate. The forecast errors are randomly distributed around the regression line and have a constant variance.
2. Not appropriate. The forecast errors are mostly positively serially correlated.
3. Very Appropriate. The errors tend first to be above the regression line and then, starting in 1997, they tend to be below the regression line until 2000 when they again are persistently above the regression line. This shows that the errors are randomly distributed.
4. Not appropriate. The forecast errors are mostly negatively serially correlated.

**76. GARCH**

**Question:**  
Which of the following features of financial asset return time-series could be captured using a standard GARCH(1,1) model (generalized autoregressive conditional heteroskedasticity)?

i) Fat tails in the return distribution  
ii) Leverage effects  
iii) Volatility clustering  
iv) Volatility affecting returns

**Pick ONE OR MORE options:**

1. (i)
2. (ii)
3. (iii)
4. (iv)