Brown University
DATA 1010
Fall 2018: Final Exam
Samuel S. Watson

Name:

You will have three hours to complete the exam, which consists of 40 questions. Among the first 36 questions, you should only solve problems for standards for which you want to improve your medal from the second exam.

No calculators or other materials are allowed, except the provided reference sheets.

You are responsible for explaining your answer to **every** *question. Your explanations do not have to be any longer than necessary to convince the reader that your answer is correct.*

For questions with a final answer box, please write your answer as clearly as possible and strictly in accordance with the format specified in the problem statement. Do not write anything else in the answer box. Your answers will be grouped by Gradescope's AI, so following these instructions will make the grading process much smoother.

I verify that I have read the instructions and will abide by the rules of the exam: ______________________________

Problem 1 **[SETFUN]**

Consider a function f from a set A to a set B. Suppose that $V \subset B$ has the property that $f^{-1}(B) = \emptyset$. Find $f^{-1}(B \setminus V)$.

Solution

Final answer:

Problem 2 **[JULIA]**

Write a Julia function which removes the vowels from a word (treating `y` as a vowel). Your function should pass the test

```
@assert remove_vowels("grasshopper") == "grsshppr"
```

Solution

Problem 3 **[LINALG]**

Consider a list of seven nonzero vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_7$, and suppose that

$$\mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3 = \mathbf{0}.$$

Suppose further that this list has a linearly independent sublist of length 6. Is it necessarily the case that $\{\mathbf{v}_4, \mathbf{v}_5, \mathbf{v}_6\}$ is linearly independent?

Solution

Problem 4 **[MATALG]**

(a) Suppose that U is a matrix with orthonormal columns. Find the matrix V such that UV is the projection matrix onto the column space of U. Write your answer in the box.

(b) Suppose that $\mathbf{v}$ is orthogonal to every column of U. Find $U'\mathbf{v}$.

Solution

Final answer:

Problem 5 **[EIGEN]**

Find a 2×2 matrix A and a unit square S with sides not parallel to the axes such that A maps S to a long skinny rectangle with sides not parallel to the axes. Feel free to specify A using an unsimplified expression.

Solution

Problem 6 **[OPT]**

Which of the following functions f has the property that the gradient descent algorithm does not necessarily converge to the global minimum of f?

$$f(x) = x(1-x) \quad g(x) = -\frac{x^3}{3} + \frac{x^2}{2} \quad h(x) = x^2(1-x)^2$$

Solution

Problem 7 **[MATDIFF]**

Suppose that A is a matrix and $\mathbf{x}$ is a vector. Differentiate $|A\mathbf{x}|^2$ with respect to $\mathbf{x}$.

Solution

Final answer:

Problem 8 **[MACHARITH]**

Explain why `a == b` returns `false` if `a = 64 - 0.5^47 - 0.5^48` and `b = 64 - 0.5^48 - 0.5^47`

Solution

Problem 9 **[NUMERROR]**

NumPy* has a function called `log1p` which returns the natural logarithm of the sum of 1 and the argument. Explain why such a function exists when you can achieve the same mathematical effect by just adding 1 and then taking the logarithm of the resulting sum. Hint: consider input values close to 0.

*R, Julia, MATLAB, Mathematica, and many other languages have such a function too.

Solution

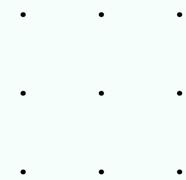
Problem 10 **[PRNG]**

Consider a pseudorandom number generator which returns integer values in the interval $[0, 2^{32} - 1]$. Is it possible for the period of the PRNG to be greater than 2^{32}?

Solution

Problem 11 **[COUNTING]**

How many triangles can be drawn with vertices at the dots shown?



Solution

Final answer:

Problem 12 **[PROBSPACE]**

Consider a random experiment which involves rolling a die, flipping a coin, and drawing a single card from a standard deck of 52 cards. Describe a probability space Ω for modeling this random experiment. How many elements does it have? Do the die and the coin correspond to disjoint subsets of Ω?

Solution

Problem 13 **[PMF]**

Suppose that a coin is weighted so that it is twice as likely to come up heads as tails. What is the probability mass function of the number of heads which appears in two independent flips of this coin?

Solution

Problem 14 **[PDF]**

Consider a ray from the origin with angle (measured with respect to the positive x-axis) chosen uniformly between 0 and 180 degrees. Let X be the x-coordinate of the point where this ray intersects the line $y = 1$. Sketch the PDF of X.

Solution

Problem 15 **[CONDPROB]**

Consider the following experiment: we roll a die, and if it shows 2 or less we select Urn A, and otherwise we select Urn B. Next, we draw a ball uniformly at random from the selected urn. Urn A contains one red and one blue ball, while urn B contains 3 blue balls and one red ball.

Find a probability space Ω which models this experiment, find a pair of events E and F such that $\mathbb{P}(E \mid F) = \frac{3}{4}$.

Solution

Problem 16 **[BAYES]**

Suppose that $Z = X + Y$ where X is 0 with probability 0.01% and 1000 with probability 99.99%, Y is a standard normal random variable, and X and Y are independent. Which of the following is closest to the conditional probability that $X = 1000$ given that $Z \geq 997$.

$$0 \quad \frac{1}{4} \quad \frac{1}{2} \quad \frac{3}{4} \quad 1$$

Solution

Final answer:

Problem 17 **[IND]**

Consider a fair die roll X. Are the events $\{X \leq 4\}$ and $\{X \in \{2, 4, 6\}\}$ independent?

Solution

Problem 18 **[EXP]**

A standard deck of cards is the Cartesian product $\{2, 3, 4, 5, 6, 7, 8, 9, 10, \mathrm{J}, \mathrm{Q}, \mathrm{K}, \mathrm{A}\} \times \{\heartsuit, \diamondsuit, \clubsuit, \spadesuit\}$. The values $\heartsuit$, $\diamondsuit$, $\clubsuit$, and $\spadesuit$ are called *suits*.

A *royal flush* is a set of five cards which are the 10, J, Q, K, and A of the same suit. The probability of a randomly chosen set of 5 cards from the deck being a royal flush is $1/649740$.

Suppose that five poker hands are dealt from a well-shuffled deck. Is it possible that all five hands are royal flushes? Find the expected number of royal flushes among the five hands.

Solution

Problem 19 **[COV]**

Can two random variables X and Y have the property that $X - Y$ has a smaller variance than both X and Y? What has to be true about the sign of $\mathrm{Cov}(X, Y)$ for this to be the case?

Solution

Problem 20 **[CONDEXP]**

Suppose that N is a geometric random variable with probability $\frac{1}{2}$, and suppose that $X_1, X_2, \ldots$ is a sequence of Poisson random variables with mean 6. Find the expected value of $X_1 + X_2 + \cdots + X_N$.

Solution

Problem 21 **[COMDISTD]**

Disprove the claim that the sum of two independent geometric random variables is a geometric random variable.

Solution

Problem 22 **[COMDISTC]**

Which distribution has smaller variance: (a) the uniform distribution on $[0, 1]$ or (b) the distribution with density proportional to e^{-8x^2}?

Solution

Final answer:

Problem 23 **[RVINEQ]**

Show that for any random variable X and for any positive u and t, we have $\mathbb{P}(X \geq u) \leq \mathrm{e}^{-tu}\mathbb{E}[\mathrm{e}^{tX}]$.

Solution

Problem 24 **[CLT]**

If $X_1, X_2, \ldots$ is a sequence of independent random variables with common density $\frac{1}{\pi(1+x^2)}$ on $\mathbb{R}$, then the distribution of $\frac{1}{n}(X_1 + X_2 \cdots + X_n)$ has the same distribution as X_1. Are the hypotheses of the central limit theorem satisfied in this case?

Solution

Final answer:

Problem 25 **[KDE]**

Discuss the behavior of the Nadaraya-Watson estimator when the bandwidth λ is very small and when λ is very large. Discuss the following phases: (1) λ is extremely small, (2) λ is not terribly small but smaller than the optimal value, (3) λ is not terribly large but larger than the optimal value, and (4) λ is extremely large.

Solution

Problem 26 **[LR]**

Explain how to find the function of the form

$$A + B \sin x + C \sin 2x + D \sin 3x$$

which has the least residual sum of squares for a given set of points $\{(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)\}$.

Solution

Problem 27 **[QDA]**

Consider applying QDA to a binary classification problem where it turns out that the plug-in estimators of the class conditional covariance matrices are exactly equal. What is the shape of the resulting decision boundary in that case?

Solution

Problem 28 **[STATLEARN]**

Consider a supervised learning model which consists of a space $\mathcal{X} \times \mathcal{Y}$, a probability measure $\mathbb{P}$, and a loss functional L. Explain the differences between the loss functional for a prediction function h and the *empirical* loss (or empirical risk) of h.

Solution

Problem 29 [NPL]

Consider a binary classification problem where the distribution of the positive class is equal to the distribution of $3 + X$ where X is an exponential random variable with parameter 1. Suppose further that the distribution of the negative class is equal to the distribution of an exponential random variable with parameter 1.

(a) What is the maximum possible detection rate?

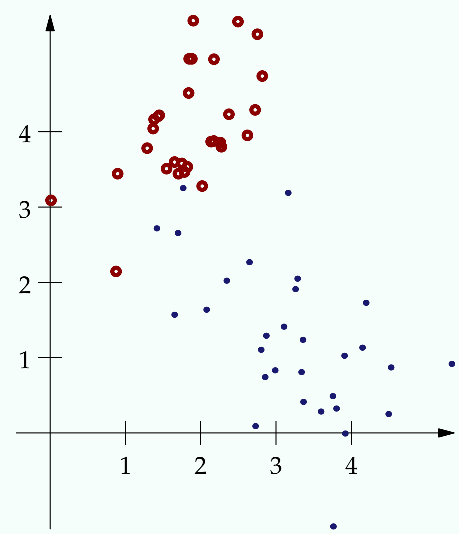
(b) What is the lowest possible false alarm rate among those prediction functions whose detection rate is equal to your answer from (a)?

Solution

Problem 30 [SVM]

(a) Find the minimum number of points that would have to be discarded to apply hard-margin SVM (directly; no kernel function) to the classification problem shown.

(b) Find a choice of decision boundary and margin width which minimizes the *number* of points with a nonzero contribution to the soft-margin SVM loss estimator. Sketch your slab and indicate the set of points with nonzero loss contribution.

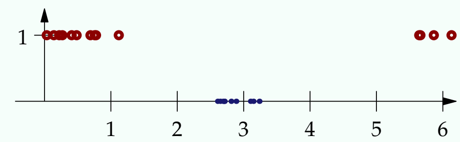


Solution

Final answer:

Problem 31 **[LOGIST]**

Consider the one-dimensional classification problem shown. The feature space is $\mathbb{R}^1$, indicated by the x-axis, and the response variable is displayed along the y-axis.



(a) Explain why logistic regression will not be able to accurately model the conditional probability of class 1 in this scenario.

(b) Sketch your best guess of the graph of the function which minimizes the logistic regression loss function. Explain your considerations for drawing the graph the way you did.

Solution

Problem 32 **[DR]**

Consider a set of 10,000 vectors in $\mathbb{R}^{784}$, each representing a handwritten digit image. Let $\mathbf{v}_k$ be the kth principal component of this set of vectors. Describe the differences between the point cloud obtained by applying $\mathbf{x} \mapsto (\mathbf{v}_1 \cdot \mathbf{x}, \mathbf{v}_{100} \cdot \mathbf{x})$ to all of the vectors and the one obtained by applying $\mathbf{x} \mapsto (\mathbf{v}_{200} \cdot \mathbf{x}, \mathbf{v}_{201} \cdot \mathbf{x})$ to them.

Solution

Problem 33 **[NN]**

Consider a neural network with ReLU activations and two affine maps: $A_1(\mathbf{x}) = \begin{bmatrix} 2 & -3 & 4 \\ 0 & -4 & 0 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 3 \\ 0 \end{bmatrix}$ and $A_2(\mathbf{x}) = \begin{bmatrix} -6 & 1 \\ 4 & 5 \\ 0 & 1 \\ 2 & -2 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 2 \\ -1 \\ 5 \\ 3 \end{bmatrix}$.

Consider a sample with $\mathbf{x}_i = \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix}$ and $y_i = \begin{bmatrix} 1 \\ -2 \\ -1 \\ 0 \end{bmatrix}$. Find the suggested change to the weight matrix W_1 assuming a learning rate of $\epsilon = 0.1$.

Solution

Final answer:

Problem 34 **[R]**

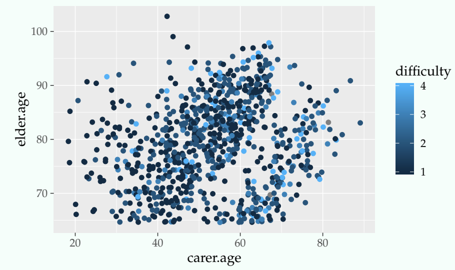
Write an R function that maps a 2×2 rotation matrix to the angle of rotation in degrees. For simplicity, you may assume that the angle is strictly between 0 and 90 degrees. The inverse tangent function in R is called `atan`, and it returns values in radians.

```
near(angle(matrix(c(cos(pi/4),sin(pi/4),-cos(pi/4),sin(pi/4)),2,2)), 45)
```

Solution

Problem 35 **[GGPLOT]**

Write `ggplot2` code to generate the plot shown for the data frame `elder.care`. Hint: the geom isn't actually `geom_point`, but rather a closely related one. Also, set the size of the points to 1.



Solution

Problem 36 **[DPLYR]**

Consider a data frame storing the standards-based medal information for a class, with column names "student.name", "type", and "standard", "medal", with an example record of "Jane Doe", "Overall", "LOGIST", "Silver", or "John Doe", "Homework", "SVM", "Bronze". Write `dplyr` code to return a data frame such that each row specifies a standard and the number of students who have an overall gold for that standard and the number of students who have an overall silver for that standard. Call the dataframe `grades`.

Solution

Problem 37 **[POINTEST]**

Suppose that T is a statistical functional, ν is a probability measure on $\mathbb{R}$, $\widehat{\theta}_1$ is a biased estimator of $\theta = T(\nu)$, and $\widehat{\theta}_2$ is a unbiased estimator of $\theta = T(\nu)$. Is it possible that the MSE of $\widehat{\theta}_1$ is smaller than the MSE of $\widehat{\theta}_2$, even though the former is biased and the latter is not? Explain why it is not possible or give an example (you do not have to do any calculations to support your example; you can just cite it).

Solution

Problem 38 **[BOOT]**

Consider the statistical functional $T(\nu)$ which returns the expectation of the minimum of two independent samples from ν.

(a) Find the *exact limiting value* of the bootstrap estimate of $T(\nu)$ given the samples

$$3, 4, 3, 7$$

from ν.

(b) Even though you found the exact value of the bootstrap estimate of $T(\nu)$, is that value necessarily close to $T(\nu)$?

Solution

Problem 39 **[HYPTEST]**

A *zero-knowledge* proof is a method by which one person can demonstrate knowledge to another person without revealing anything beyond the fact of that knowledge.

For example, consider a red ball, a green ball, and a color blind friend who is skeptical that the balls are actually distinguishable. It is possible to convince your friend that you can distinguish the balls without revealing which ball is red and which is green: have the friend hold both balls behind his back and choose one of them randomly to reveal to you. Then he puts both balls behind his back again and reveals a ball for a second time, switching them behind his back with probability $1/2$. You then indicate whether he switched the balls.

Describe this random experiment in a hypothesis test framework. Identify the null hypothesis and the alternative hypothesis. Determine the number of times the experiment must be repeated to reject the null hypothesis with 99% confidence (bearing in mind that the balls are actually different colors, so you will in fact be able to distinguish them every time).

Solution

Problem 40 **[MLE]**

Consider a Bernoulli distribution with unknown parameter p. Show that the maximum likelihood estimator of p is equal to the proportion of 1's.

Solution