1.

Summary of main points

The paper discusses the culture prevalent in the statistics community that we assume a model to the data prior to studying and analyzing a data set. Assuming the mechanism of the data's nature leads to bad conclusions, and quells innovation. A much better approach is to make no assumptions about the mechanism or nature of the data. An unbiased observation of the data allows for the construction of accurate data models. In other words, we build the algorithms which accurately represent the data instead of forcing the data to fit in a predefined model. One key note from the paper is to obtain domain knowledge of the data before constructing data models. This is not always viable however as the data can be too complex for humans to obtain a domain knowledge or at least an accurate domain knowledge. A lack of domain knowledge hurts our ability as humans to build appropriate models. Neural Networks approximate an algorithm for the data which improves with training. The training allows for the study of complex and large data which might not fit into any of the standard statistical models in use today. This removes the requirement to obtain domain knowledge as the domain knowledge is integrated into the trained models.

2.

A.

A matrix U is orthogonal if U*UT is the identity matrix. The euclidean norm of U is therefore an identity matrix yielding the euclidean norm of any matrix X times the orthogonal matrix to be itself.

B.

Orthogonal matrices are rotation matrices with the requirement that the rows and columns are independent. Given the identity matrix, if we rotate it, we are given the two listed matrices. This is also the reason that finding the magnitude

3.

A.

I. If you are given Y, then the expectation of Y is guaranteed. Therefore the only condition to determine is the expectation of X. You could even chain the expectation of Z and append a given z with no change.

II. Lets say that C contains 10 items and the population outside of C is 5 items. The probability of C is 10/15. You should expect a 1 if C is chosen and a 0 otherwise. It makes no sense to expect any more or less than 10/15 for C to be chosen. The expectation of an event is equivalent to the probability of the event.

III. Independent events do not effect each other in terms of probability. And since probability equates to expectation, independent events do not effect each other in terms of expectation. Therefore the expectation of the two events together is equivalent to the expectation of the two events individually.

B.

I. <= There is no such thing as a negative probability. With this in mind 0 to 1 * x is at most equal to x, but possibly less than x.

II. = If Y is given, then we only care about the probability of X to give the probability of x. Because they are independent, we can disregard y in this case.

III. >= X AND Y >= Y AND X AND X --> 1 >= X as X is from 0 to 1.

4. ATTACHED AT END

D. A matrix is PD iff all eigenvalues are positive. If you sum two matrices that are both PD, then the resultant eigenvalues will also be positive.

E. For PD, x > 0 so PD+PSD means that for every x, x is now > 0 even if it wasn't prior which is the definition of PD.

F.

i. Neither as it is not a square symmetric matrix.

ii. The eigenvalues are positive, positive so it is PD

iii. The eigenvalues are 0, positive, positive therefore PSD.

iv. The eigenvalues are all 1, therefore PD

v. neither as all eigenvalues are -1

vi. All eigenvalues are positive, therefore PD.

vii. All eigenvalues are positive, therefore PD.

viii. All eigenvalues are positive, therefore PD.

5.

A. F1

GRADIENT:

fv1(v1,v2) = 2*v1 + 3*e^v2

fv2(v1,v2) = 3*v1*e^v2

HESSIAN:

fv1v1(v1,v2) = 2

fv1v2(v1,v2) = 3*e^v2

fv2v1(v1,v2) = 3*e^v2

fv2v2(v1,v2) = 3*v1*e^v2

B. F2

GRADIENT:

fv1(v1,v2) = 12*v2*v1^2 - v2*log(v2)

fv2(v1,v2) = 4*v1^3 - v1 - v1/v2

HESSIAN:

fv1v1(v1,v2) = 24*v2*v1

fv1v2(v1,v2) = 12*v2 - 1 - 1/v2

fv2v1(v1,v2) = 3 - 1/v2

fv2v2(v1,v2) = 1 / v2^2

C.

JACOBIAN:

[2*v1 + 3*e^v2	12*v2*v1^2 - v2*log(v2)
3*v1*e^v2	4*v1^3 - v1 - v1/v2]

6. ATTACHED AT END

4. A. $A = U \Lambda U^T$

$Ax = \lambda x$

$x^T A x = \lambda x^T x$

$(x^T A x)^T = \bar{\lambda} x^T x$

$A = x \bar{\lambda} x^T = U \Lambda U^T$

B. $x^T A x \geq 0$

$\lambda_i \geq 0$

From problem 4A:

$\lambda x^T x = x^T A x$

$\lambda = \frac{x^T A x}{x^T x} = \frac{x^T A x}{\|x\|^2} \geq 0$

C. $x^T A x > 0$

$\lambda_i > 0$

$\lambda x^T x = x^T A x$

$\lambda = \frac{x^T A x}{\|x\|^2} > 0$

6 A: Calculate Bias of $\bar{X}$

$\text{Bias} = E[\hat{\theta}] - \theta$

$E[\bar{x}] = E[\frac{1}{n} \Sigma x_i]$

$= \frac{1}{n} E[\Sigma x_i]$

$= \frac{n}{n} E[x_1]$

$= \frac{n\mu}{n} =$

$\text{Bias} = \frac{n\mu}{n} - \mu = 0$

B: Find $Var[\bar{x}]$

$Var[\bar{x}] = Var[\frac{1}{n} \Sigma x_i]$

$= \frac{1}{n^2} Var[\Sigma x_i]$

$= \frac{1}{n} Var[x_i]$

$= \frac{\theta^2}{n}$

C. $MSE[\bar{x}] = \text{Bias}^2 + \text{Variance}$

$\text{Bias}^2 = 0$

$MSE = + \frac{\theta^2}{n}$

D. $\text{Bias}[\hat{s}^2] = E[\hat{s}^2] - \theta^2$

$E[\hat{s}^2] = E[\frac{1}{n} \Sigma (x_i - \bar{x})^2]$

$= \frac{1}{n} E[\Sigma (x_i - \bar{x})^2]$ // Factor out $\bar{X}$

$= \frac{1}{n} E[\Sigma x_i^2 - n\bar{x}^2]$

$= \frac{1}{n} (n(\mu^2 + \theta^2) - n(\mu^2 + \frac{\theta^2}{n}))$

$= \mu^2 + \theta^2 - \mu^2 - \frac{\theta^2}{n}$

$= \theta^2 - \frac{\theta^2}{n}$

$\text{Bias} = -\frac{\theta^2}{n}$