Submission Deadline : 12/9/2021 20:00

Mid Sem Exam

Q.1 Assume that $X_1, X_2, \cdots X_n$ are i.i.d. sampled from a Gaussian likelihood $N(\mu, \sigma^2)$.

(a) What will be the likelihood of the sequence of observations

$f(x_1, x_2, \cdots x_n | \mu, \sigma)$

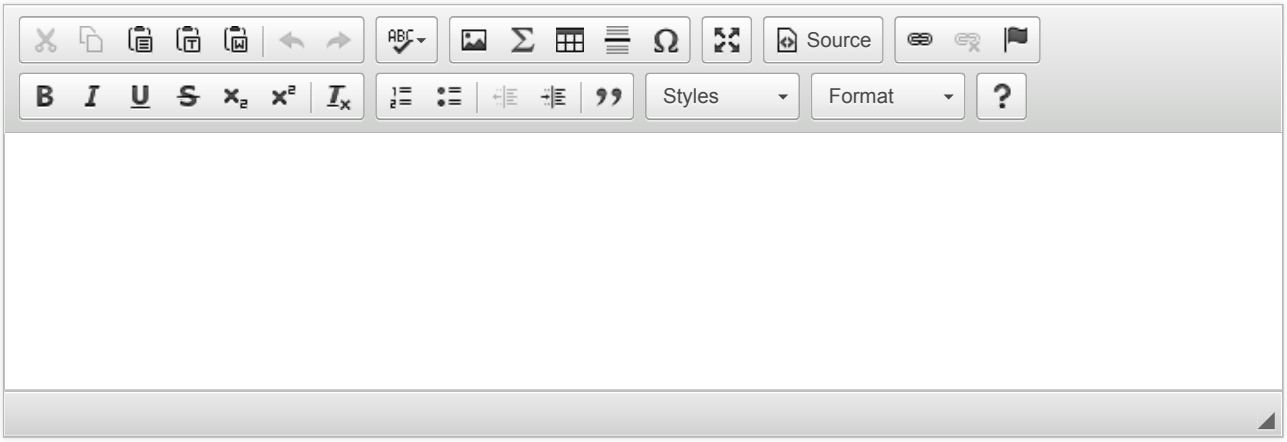
(b) Can you calculate the MLEs for μ and σ?

Max. score: 8; Neg. score: 0

This is a long answer type question. You can either upload a file or type your answer below.

UPLOAD A FILE

or



Q.2

I am giving you some incomplete code for gradient descent-based optimization. Here, *gradient* is a function passed as a parameter to the gradient_descent function. *Gradient* specifies the numerical value of the gradient at any x value passed to it.

```
def gradient_descent(gradient,init_,learn_rate, n_iter=50, tol=1e-06):
  x = init_
  for _ in range(n_iter):
    delta = _____________
    if np.all(np.abs(delta) <= tol):
      break
    x += _______
  return x
```

See below for an example of how to use this function to find the global minimum for the function $y = 2x^2 + 3x + 5$

```
gradient_descent(gradient=lambda v: 4 * v + 3, init_=4.0, learn_rate=0.2)
```

(a) Complete the function [2 marks]

(c) Can you modify the gradient descent algorithm I have defined above to create a stochastic gradient descent algorithm, with mini-batch capabilities? Write your answer as pseudo-code using all the variables defined in the code I have given you, plus whatever variables you need to define. [6 marks]

Max. score: 12; Neg. score: 0

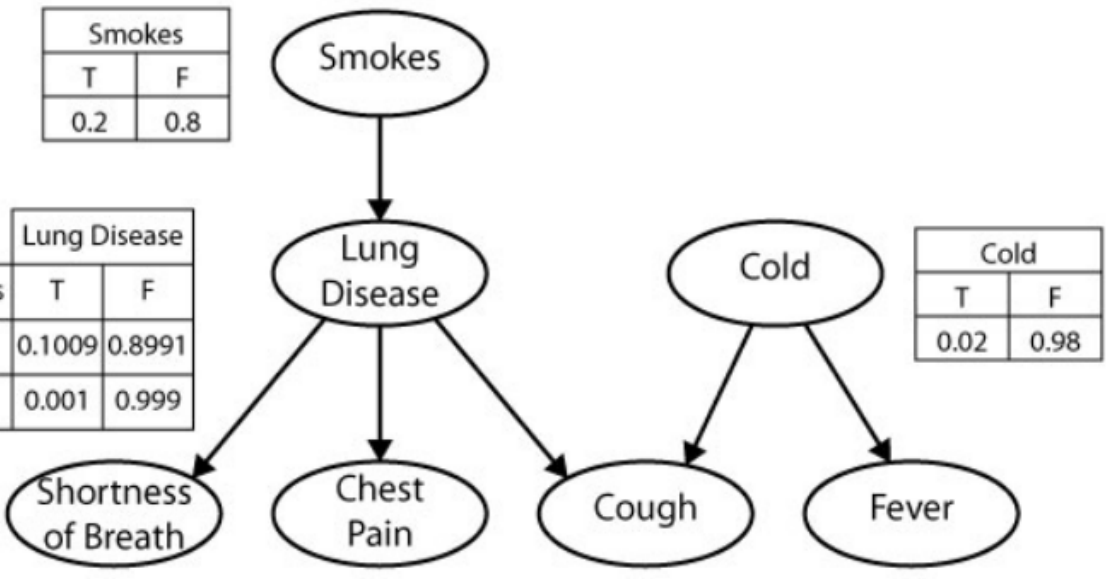
This is a long answer type question. You can either upload a file or type your answer below.

UPLOAD A FILE

or

Source Styles Format

Q.3



Smokes	
T	F
0.2	0.8

Smokes	Lung Disease T	Lung Disease F
T	0.1009	0.8991
F	0.001	0.999

Cold	
T	F
0.02	0.98

Lung Disease	Shortness of Breath T	Shortness of Breath F
T	0.208	0.792

Lung Disease	Chest Pain T	Chest Pain F
T	0.208	0.792

Lung Disease	Cold	Cough T	Cough F
T	T	0.7525	0.2475
T	F	0.505	0.495

Cold	Fever T	Fever F
T	0.307	0.693
F	0.01	0.99

I have specified a Bayesian network above using conditional probability tables. For conditional probability tables in this format – the conditioning variable values are listed along the rows on the left, the variable being conditioned is listed up top. The direction of arrows in the network specifies our understanding of causality in this system.

Calculate the following probabilities. Approximating to the first decimal place is fine, even in intermediate steps. Leaving the numbers uncalculated in the formula is also acceptable, but be aware that doing things this way will increase the chances of making mistakes.

The probability that someone may have a cold [1 mark]

The probability that someone may have lung disease [2 marks]

The probability that someone may have shortness of breath [2 marks]

The probability that someone who has a fever may have a cold [3 marks]

Max. score: 8; Neg. score: 0

This is a long answer type question. You can either upload a file or type your answer below.

UPLOAD A FILE

or

Source Styles Format

Q.4 Let C be a convex set and $\lambda_1, \lambda_2, \cdots, \lambda_p \geq 0$ such that $\sum_{i=1}^{p} \lambda_i = 1$. If $a_1, a_2, \cdots a_p \in C$ then prove that $\sum_{i=1}^{p} \lambda_i a_i \in C$

Max. score: 5; Neg. score: 0

This is a long answer type question. You can either upload a file or type your answer below.

UPLOAD A FILE

or

Source Styles Format

Q.5 I took a standard deck of playing cards and replaced the queen of spades with an extra queen of diamonds. Now, I draw a card from this deck and consider the following three random events
X = the card is a queen
Y = the card is red
Z = the card is hearts

Can you tell me which pairs of the random variables XY, YZ and ZX are independent and why?

Max. score: 5; Neg. score: 0

This is a long answer type question. You can either upload a file or type your answer below.

UPLOAD A FILE

or

Source Styles Format

Q.6 When minimizing a convex function f using (sub)gradient descent (without any constraints), step lengths don't matter since f is convex

Max. score: 1; Neg. score: 0

[Imp. Note: If you wish to skip the question, you should do so immediately. Once an option is chosen (either true or false), you can not skip the question at a later stage.]

- false
- true

Q.7 Suppose X is a random variable such that $x \geq 1 \; \forall x \in s(X)$ where s(X) is the support of X. Then the variance of X must be greater than 1 too i.e. $Var[X] \geq 1$

Max. score: 1; Neg. score: 0

[Imp. Note: If you wish to skip the question, you should do so immediately. Once an option is chosen (either true or false), you can not skip

Q.8 Let's say we have two random variables X and Y, s.t. X + Y = 2. If E(X) = 3, then E(Y) has to be equal to -1.

Max. score: 1; Neg. score: 0

[Imp. Note: If you wish to skip the question, you should do so immediately. Once an option is chosen (either true or false), you can not skip the question at a later stage.]

- false
- true

Q.9 A twice differentiable function f is convex if and only if it's Hessian is positive semi-definite at every point in its domain

Max. score: 1; Neg. score: 0

[Imp. Note: If you wish to skip the question, you should do so immediately. Once an option is chosen (either true or false), you can not skip the question at a later stage.]

- false
- true

Q.10 It is impossible for the decision boundary of a k-nearest neighbors classifier to be linear.

Max. score: 1; Neg. score: 0

[Imp. Note: If you wish to skip the question, you should do so immediately. Once an option is chosen (either true or false), you can not skip the question at a later stage.]

- false
- true

Q.11

$\max_{x,y} x^2 + y^2$

$x \in [1, 2]$

$y \in [1, 5]$

The global minimum of this objective function is 3

Max. score: 1; Neg. score: 0

[Imp. Note: If you wish to skip the question, you should do so immediately. Once an option is chosen (either true or false), you can not skip the question at a later stage.]

- false

Q.12 If X and Y are two random variables with cov(X,Y) = 0, then they are independent random variables

Max. score: 1; Neg. score: 0

[Imp. Note: If you wish to skip the question, you should do so immediately. Once an option is chosen (either true or false), you can not skip the question at a later stage.]

- false
- true

Q.13 To avoid over-fitting in decision trees, we must be sure to use information gain as the criterion to split nodes.

Max. score: 1; Neg. score: 0

[Imp. Note: If you wish to skip the question, you should do so immediately. Once an option is chosen (either true or false), you can not skip the question at a later stage.]

- false
- true

Q.14 Using an L1 norm in the regression loss function ensures that our model ignores statistical outliers in the data features

Max. score: 1; Neg. score: 0

[Imp. Note: If you wish to skip the question, you should do so immediately. Once an option is chosen (either true or false), you can not skip the question at a later stage.]

- false
- true

Q.15 The computational complexity of the prediction step in linear regression is independent of the size of the training dataset.

Max. score: 1; Neg. score: 0

[Imp. Note: If you wish to skip the question, you should do so immediately. Once an option is chosen (either true or false), you can not skip the question at a later stage.]

- false
- true

Q.16 Sub-gradient descent can be used to find global optima for any loss function – differentiable or non-differentiable

- false
- true

Q.17 The reason we use minibatches in SGD is that a single training example is usually not a great guide for finding the gradient of the loss function

Max. score: 1; Neg. score: 0

[Imp. Note: If you wish to skip the question, you should do so immediately. Once an option is chosen (either true or false), you can not skip the question at a later stage.]

- false
- true

SAVE SUBMIT