2

JOMO KENYATTA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY

CONTINUOUS ASSESSMENT TEST 2019/2020

STA 2408: REGRESSION MODELLING
CAT

INSTRUCTIONS TO CANDIDATES:

* *Answer question all the questions*
* *Be neat and show all your workings*

---

1. Differentiate between parametric and non parametric regression [2 marks]

2. Define what is meant by a spline and illustrate using two examples. Show that splines are linear smoothers.

3. Draw a well labelled diagram of a neural network with two hidden layers and X1; X2 and X3 as the input variables.

4. Define stochastic and deterministic designs under Kernel regression estimation stating their mean functions.

5. K-NN is a weighted average in a varying neighborhood(defined through those X variables which are among the K nearest neighbors of X in Euclidean distance (linear real distance)). Let $(X_i, Y_i)_{i=1}^{5}$ be $(1,5),(7,12),(3,1),(2,0),(5,4)$ Compute (K-NN estimate) $\hat{m}_k(x)$ (i.e) for $x = 4 and k = 3$.

~~X~~6. Show that a linear regression estimator is a special case of non parametric estimator.

7. Given the data

X	20	30	15	25	28
Y	45.6	35.3	40.3	20.0	43.2

where X and Y are predictor and response variables,respectively.Using a rectangular Kernel function and 3-nearest neighbour(K-NN)find the estimate of

i) Density function of X at point x=29 [3 marks]

ii) regression function at point x=29 [3 marks]

8. There is a functional relationship between the mass density p of air and the altitude h above the sea level.

Altitude above the sea level, h (km)	0.32	0.64	1.28	1.60
Mass Density, $\rho (kg/m^3)$	1.15	1.10	1.05	0.95

In the regression model $\rho = k_1 e^{-k_2 h}$, the constant k_2 is found as $k_2 = 0.1315$. assuming that the mass density of air at the top of the atmosphere is $\frac{1}{1000^{th}}$ of the mass density of air at the sea level.
Calculate the altitude in kilometers of the top of the atmosphere. [8 marks]

$$\rho = \frac{\sum_{i=1}^{4} k_i e^{-k_2 h_i}}{\sum_{i=1}^{4} e^{-k_2 h_i}}$$

0.000208 < 0.2838

1. Differentiate between the following models

   (a) Linear and non-linear models [2 marks]

   (b) Parametric and non-parametric models [2 marks]

2. For an exponential model $y = \gamma e^{\zeta x}$ that is best fit to the data $(x_1, y_1), ..., (x_n, y_n)$ derive a non-linear equation that can be used to estimate the value ζ. [5 marks]

3. Empirical results have shown that the rate of gas flow from a container is proportional to some power of the nozzle pressure. Below is given the flow rate, in cm^3 per second as a function of pressure.

y Flow rate, F (cm^3/sec)	88	134	135	148	172	240
x Pressure, P (psi)	15	22	26	28	52	60

   The rate of gas flow is related to the nozzle pressure via the regression model $F = \alpha e^{P\beta}$. By transforming the above data, find

   (a) The value of the regression parameters α and β [6 marks]

   (b) The rate of gas flow if the nozzle pressure is increased to 80 psi [2 marks]

4. The following data contains measurements of yield from an experiment done at six different temperature levels.

Temperature (T)	Yield (Y)
5	6.26
10	4.24
15	3.88
20	2.26
25	1.44
30	0.60

   If the Yield behaves as a second order polynomial function of temperature fit the above data to the model $Y = a_0 + a_1 T + a_2 T^2$

   Y

   [7 marks]

5. There exists a functional relationship between the mass density ρ of air and the altitude h above the sea level. A sample data of the two variables is given below

Altitude above sea level, h(km)	0.32	0.64	1.28	1.60
Mass Density, ρ (kg/m^3)	1.15	1.10	1.05	0.95

   The functional relationship can be expressed using the regression model $\rho = k_1 e^{-k_2 h}$, where the constant k_2 is found as $k_2 = 0.1315$. Assuming that the mass density of air at the top of the atmosphere is $1/1000^{th}$ of the mass density of air at sea level find the altitude in kilometers of the top of the atmosphere. [6 marks]

0.0012

JOMO KENYATTA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY

CONTINUOUS ASSESSMENT TEST 2019/2020

STA 2408: REGRESSION MODELLING

CAT

INSTRUCTIONS TO CANDIDATES:

* *Answer question all the questions*

* *Be neat and show all your workings*

---

1. Differentiate between parametric and non parametric regression [2 marks]

2. Define the term smoothing with respect to non parametric regression. [2 marks]

3. What is Kernel regression. Illustrate your explanation with examples. [4mks]

4. State the steps involved in fitting a Local Polynomial Regression Curve [3mks]

5. K-NN is a weighted average in a varying neighborhood(defined through those X variables which are among the K nearest neighbors of X in Euclidean distance (linear real distance)). Let $(X_i, Y_i)_{i=1}^{5}$ be $(1,5), (7,12), (3,1), (2,0), (5,4)$ Compute (K-NN estimate) $\hat{m}_k(x)$ (i.e) for $x = 4 and k = 3$.

6. Many patients get concerned when a test involves injection of a radioactive material.For example for scanning a gallbladder, a few drops of Technetium-99m isotope is used. Half of the technetium-99m would be gone in about 6 hours. It, however, takes about 24 hours for the radiation levels to reach what we are exposed to in day-to-day activities. Below is given the relative intensity of radiation as a function of time.

t (hrs)	0	1	3	5	7	9
ρ	1.000	0.891	0.708	0.562	0.447	0.355

In the regression model $\rho = Ae^{\lambda t}$, the constant λ is found as $k_2 = 0.11505$. Find the value of the constant term A, the half-life of Technium-99m, and the radiation intensity after 24 hours.

7. For a given sample $(X_i, Y_i), i = 1, 2, \cdots n$ from an NPR model $Y_i = m(X_i) + \epsilon_i$, where $n = 140$, $X_i \in [0.1894, 2.6855]$ and ϵ_i are i.i.d with $E\epsilon_1^2 = \sigma^2$, we define $s_n(x) = \sum_{i=1}^{n} K_h(X_i - x)(X_i - x)^l$ and $T_{nl}(x) = \sum_{i=1}^{n} K_h(X_i - x)(X_i - x)^l Y_i$. We used the standard Gaussian kernel and $h = 0.2420$. Let $x_0 = 1$. We obtained $s_{nl}(x_0), l =, 0, 1, 2$ as $22.8038, -0.5758, 1.1189$ respectively, and $T_{nl}(x_0), l = 0, 1$ as $24.0092, 0.1631$ respectively. Compute the local linear estimators of $m(x_0)$ and $m'(x_0)$ respectively.

1. Differentiate between the following models
   (a) Linear and non-linear models [2 marks]
   (b) Parametric and non-parametric models [2 marks]
2. For an exponential model $y = \gamma e^{\zeta x}$ that is best fit to the data $(x_1, y_1), ..., (x_n, y_n)$ derive a non-linear equation that can be used to estimate the value ζ. [5 marks]
3. Empirical results have shown that the rate of gas flow from a container is proportional to some power of the nozzle pressure. Below is given the flow rate, in cm^3 per second as a function of pressure.

Flow rate, F (cm^3/sec)	88	134	135	148	172	240
Pressure, P (psi)	15	22	26	28	52	60

The rate of gas flow is related to the nozzle pressure via the regression model $F = \alpha e^{P\beta}$. By transforming the above data, find
   (a) The value of the regression parameters α and β [6 marks]
   (b) The rate of gas flow if the nozzle pressure is increased to 80 psi [2 marks]
4. The following data contains measurements of yield from an experiment done at six different temperature levels.

Temperature (T)	Yield (Y)
5	6.26
10	4.24
15	3.88
20	2.26
25	1.44
30	0.60

If the Yield behaves as a second order polynomial function of temperature fit the above data to the model $Y = a_0 + a_1 T + a_2 T^2$

[7 marks]
5. There exists a functional relationship between the mass density ρ of air and the altitude h above the sea level. A sample data of the two variables is given below

Altitude above sea level, h(km)	0.32	0.64	1.28	1.60
Mass Density, ρ (kg/m^3)	1.15	1.10	1.05	0.95

The functional relationship can be expressed using the regression model $\rho = k_1 e^{-k_2 h}$, where the constant k_2 is found as $k_2 = 0.1315$. Assuming that the mass density of air at the top of the atmosphere is $1/1000^{th}$ of the mass density of air at sea level find the altitude in kilometers of the top of the atmosphere. [6 marks]

9.05 9.50 57