Name: Matric No.:

Instructions: Please write your name and matric. no. on every sheet. State your assumptions, if any.

1. (0.5 points) 6LoWPAN is used in IoT networks for which of the following reasons?
   a) To allow IEEE 802.15.4 networks to send data using IPv6
   b) To solve security issues
   c) To provide device localization
   d) To provide reliability against packet loss

   Choice: a. — No explanation/justification is needed.

2. (0.5 points) The value proposition of an IoT based system is in the use of the data that it generates.
   a) True
   b) False

   Choice: a — No explanation/justification is needed.

3. (0.5 points) SIGFOX and LoRa are technologies that compete with WiFi in providing high data rate network access to IoT devices.
   a) True
   b) False

   Choice: b — No explanation/justification is needed.

4. (0.5 points) The maximum likelihood estimator is asymptotically unbiased.
   a) True
   b) False

   Choice: a — No explanation/justification is needed.

5. (0.5 points) Consider a linear regression model that perfectly fits the training data (training error is zero). Then, which of the following statements in true?
   a) The error on test data will always be zero.
   b) You can never have zero error on test data.
   c) None of the above.

   Choice: c — No explanation/justification is needed.

6. (0.5 points) Naive Bayes cannot capture interdependencies between variables.
   a) True
   b) False

   Choice: a — No explanation/justification is needed.

7. Consider a Naive Bayes classifier. The data belongs to two classes. We denote the class as y and it is known that $P[y=0]=0.5$ and $P[y=1]=0.5$. The input data has 50 feature dimensions and each of these are represented as $x_1, x_2, \cdots, x_{50}$. The features are all binary, i.e., they can only take values of 0 and 1. Also, all the features have the same conditional probability:

$$\begin{aligned} P[x_i=1|y=0] &= a, \qquad 1 \le i \le 50 \\ P[x_i=1|y=1] &= b, \qquad 1 \le i \le 50 \end{aligned} \tag{1}$$

   a) (2 points) Consider a data sample with alternating feature values: $X=(x_1=1, x_2=0, x_3=1, x_4=0, \cdots, x_{50}=0)$. Compute $P[y=1|X]$.

$$P[X|y=0] = \prod_{i=1}^{50} P[x_i|y=0] = a^{25}(1-a)^{25}$$

$$P[X|y=1] = \prod_{i=1}^{50} P[x_i|y=1] = b^{25}(1-b)^{25}$$

$$P[y=1|X] = \frac{P[X|y=1]\,P[y=1]}{P[X]} = \frac{P[X|y=1]\,P[y=1]}{P[X|y=1]\,P[y=1]+P[X|y=0]\,P[y=0]} = \frac{b^{25}(1-b)^{25}}{b^{25}(1-b)^{25}+a^{25}(1-a)^{25}}$$

   b) (1 point) Find the class that the data sample in a) belongs to if $a=0.4$ and $b=0.3$.

$$\text{class} = \arg\max_y P[X|y]\,P[y]$$

$$P[X|y=0]\,P[y=0] = a^{25}(1-a)^{25} \times 0.5 = 0.16\times 10^{-15}$$

$$P[X|y=1]\,P[y=1] = b^{25}(1-b)^{25} \times 0.5 = 0.57\times 10^{-17}$$

$\Rightarrow$ class $y=0$

Name: Matric No.:

8. Research has shown that high humidity in the environment is the dominant factor that leads to failures in disk drives. A manufacturer of disk drives tests some of its disks in a laboratory environment with a high humidity level (80% relative humidity). The manufacturer tests n hard disks in this condition and the result for each disk is denoted by x_i, $1 \leq i \leq n$. The result of each test is marked as 0 for "pass" and 1 for "fail". You can consider each test to be independent of other tests.

   a) (1 point) What would be a reasonable probability distribution to model the probability that a hard disk fails under these humidity conditions? Justify your answer in no more than two sentences.

   Each test is independent and output is binary $\Rightarrow$ Bernoulli

   b) (2 points) Let p denote the probability that a hard disk fails in these conditions. Use the method of moments to estimate p. Your answer should be in terms of $x_1, x_2, \cdots, x_n$ and n.

   For a Bernoulli RV: $E[X] = p$

   Equating the theoretical first moment to sample first moment: $p = \frac{1}{n}\sum_{i=1}^{n} x_i$

   $\Rightarrow \hat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i$

   c) (2 points) Find the maximum likelihood estimator for p. The answer should be in terms of $x_1, x_2, \cdots, x_n$ and n.

   Likelihood function: $L(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$

   log-likelihood function: $l(p) = \log p \sum_{i=1}^{n} x_i + \log(1-p) \sum_{i=1}^{n}(1-x_i)$

   $\frac{d\,l(p)}{dp} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\sum_{i=1}^{n}(1-x_i)}{1-p} \overset{\text{set}}{=} 0 \Rightarrow \hat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i$

9. Consider the following data regarding an independent variable, x, and a dependent variable, y:

   a) (2 points) Fit a linear regression model to the data.

xy	x^2	x	y
4.80	1.44	1.20	4.00
12.88	5.29	2.30	5.60
24.49	9.61	3.10	7.90
27.20	11.56	3.40	8.00
40.40	16.00	4.00	10.10
47.84	21.16	4.60	10.40
66.00	30.25	5.50	12.00
223.61	95.31	24.10	58.00

   $\beta_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{7 \times 223.61 - 24.1 \times 58}{7 \times 95.31 - 24.1^2}$

   $= 1.94$

   $\beta_0 = \frac{\sum y - \beta_1 \sum x}{n} = \frac{58 - 1.94 \times 24.1}{7}$

   $= 1.61$

   b) (2 points) Is a linear regression model a good choice for this data? Justify your answer.

   $R^2 = 1 - \frac{SSE}{SST} = 0.98$

   Since R^2 is very close to 1, we conclude it is a good model.

x	y	$\hat{y}$	SSE	SST
1.20	4.00	3.94	0.004	18.367
2.30	5.60	6.07	0.221	7.213
3.10	7.90	7.62	0.078	0.149
3.40	8.00	8.21	0.044	0.082
4.00	10.10	9.37	0.533	3.252
4.60	10.40	10.53	0.017	4.470
5.50	12.00	12.28	0.078	13.796
			0.975	47.369