Birla Institute of Technology & Science, Pilani
Work Integrated Learning Programmes Division
Second Semester 2018-2019
M.Tech (Data Science and Engineering)
Mid-Semester Test (EC-2 Regular)

SOLUTION

Course No.	: DSECL ZG565	No. of Pages = 2
Course Title	: MACHINE LEARNING	No. of Questions = 6
Nature of Exam	: Closed Book	
Weightage	: 30%	
Duration	: 90 minutes	
Date of Exam	: August 11, 2019 (FN)	

Note:
1. Please follow all the *Instructions to Candidates* given on the cover page of the answer book.
2. All parts of a question should be answered consecutively. Each answer should start from a fresh page.
3. Assumptions made if any, should be stated clearly at the beginning of your answer.

Answer All the Questions (only on the pages mentioned against questions. if you need more pages, continue remaining answers from page 20 onwards)

Question 1. [Marks 3+1=4] **[to be answered only on pages 3-5]**

Derive the equation and shape of decision surface for real-valued random variable **X**=<X_1, X_2, .., X_n> and boolean output Y for logistic regression. P(Y=1|**X**) is given by

$$P(Y=1|X=<X_1,...X_n>) = \frac{1}{1+exp(w_0+\sum_i w_iX_i)}$$

$$P(Y=1|X=<X_1,...X_n>) = \frac{1}{1+exp(w_0+\sum_i w_iX_i)}$$

implies

$$P(Y=0|X=<X_1,...X_n>) = \frac{exp(w_0+\sum_i w_iX_i)}{1+exp(w_0+\sum_i w_iX_i)}$$

implies

$$\frac{P(Y=0|X)}{P(Y=1|X)} = exp(w_0+\sum_i w_iX_i)$$

Thus decision surface is given by

$$\ln\frac{P(Y=0|X)}{P(Y=1|X)} = w_0+\sum_i w_iX_i$$

Shape of decision surface is a hyperplane (straight line in 2 dimension)

Question 2. [Marks 4+1=5] **[to be answered only on pages 6-7]**

Consider the hypothesis function $h(\mathbf{x}) = w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1^2 + w_4 x_2^2$ with learnt $\mathbf{w} =< w_0, w_1, w_2, w_3, w_4 > = < -36, 0, 0, 4, 9 >$. What is the equation and shape of the decision boundary $g(x_1, x_2)$ for logistic regression given by $P(Y = 1|\mathbf{X}) = \frac{1}{1+e^{h(\mathbf{x})}}$

Decision boundary $g(x_1, x_2)$ is given by

$$\ln \frac{P(Y=0|X)}{P(Y=1|X)} = w_0 + \sum_i w_i X_i$$

$$\text{Or, } g(x_1, x_2) = -36 + 4x_1^2 + 9x_2^2 = -1 + \frac{x_1^2}{9} + \frac{x_2^2}{4} = 0$$

Shape of the decision surface is ellipse centered at (0,0) and major axis aligned to x_1.

If $P(Y = 1|\mathbf{X}) = \frac{1}{1+e^{-h(\mathbf{x})}}$ is used, the decision surface remains the same. Only the class of training data points inside the ellipse reverses.

Question 3. [Marks 1+1+2.5+2.5+1=8] **[to be answered only on pages 8-10]**

First five documents in the following figure are used to train a Naive Bayes classifier. Calculate Prob(+), Prob(-), Prob(+ | Test), Prob(- | Test) for the bag of words model. Which class does the Test document belong to?

	Cat	Documents
Training	-	just plain boring
	-	entirely predictable and lacks energy
	-	no surprises and very few laughs
	+	very powerful
	+	the most fun film of the summer
Test	?	predictable with no originality

$$P(-) = \frac{3}{5} \qquad P(+) = \frac{2}{5}$$

The likelihoods from the training set for the four words "predictable", "with", "no", and "originality", are as follows,

$$P(\text{"predictable"}|-) = \frac{1+1}{14+20} \qquad P(\text{"predictable"}|+) = \frac{0+1}{9+20}$$

$$P(\text{"with"}|-) = \frac{0+1}{14+20} \qquad P(\text{"with"}|+) = \frac{0+1}{9+20}$$

$$P(\text{"no"}|-) = \frac{1+1}{14+20} \qquad P(\text{"no"}|+) = \frac{0+1}{9+20}$$

$$P(\text{"originality"}|-) = \frac{0+1}{14+20} \qquad P(\text{"originality"}|+) = \frac{0+1}{9+20}$$

For the test sentence S = "predictable with no originality", the chosen class, is therefore computed as follows:

$$P(S|-)P(-) = \frac{3}{5} \times \frac{2 \times 1 \times 2 \times 1}{34^4} = 1.8 \times 10^{-6}$$

$$P(S|+)P(+) = \frac{2}{5} \times \frac{1 \times 1 \times 1 \times 1}{29^4} = 5.7 \times 10^{-7}$$

The model thus predicts the class *negative* for the test sentence.

The above solution assumes Laplacian smoothing.

Question 4. [Marks 5] **[to be answered only on page 11]**

a) Consider a classification model with logistic regression and L2 regularization. Assuming that model is suffering from the problem of over-fitting, decreasing the value of regularization parameter helps in reduction of over-fitting. **False**

b) In the case of large feature space, Naïve Bayes algorithm outperforms logistic regression. **True** (NB error is bounded by O(logN) viz-a-viz O(N) for LR, for dimension N)

c) Gaussian Naive Bayes classifier can have linear decision surface. **True** (when training data distribution for different classes are mean-shifted but same, decision boundary is a hyderplane)

d) Bagging is used in decision tree to reduce bias. **False** (bagging reduces overfitting but not bias, which is why feature randomization performed in random forest to reduce bias)

e) What techniques can be used to reduce overfitting in Decision tree? i) **Pruning** ii) **Bagging**

Question 5. [Marks 1+3=4] **[to be answered only on pages 12-14]**

a) A coin is tossed 250 times and lands heads 50 times. What is the maximum likelihood estimate for θ = probability of heads?
$\theta = 50 / 250 = 1/5$

b) A 6-sided die is rolled 16 times resulting in 2 ones, 4 twos, 0 threes, 5 fours, 2 fives, 3 sixes. What is the maximum likelihood estimate for all values of θ_i where i is <1,2,3,4,5,6> for each side of the die?
$\boldsymbol{\theta} = < \theta_1, \theta_2, .., \theta_6 > = < 2/16, 4/16, 0/16, 5/16, 2/16, 3/16>$

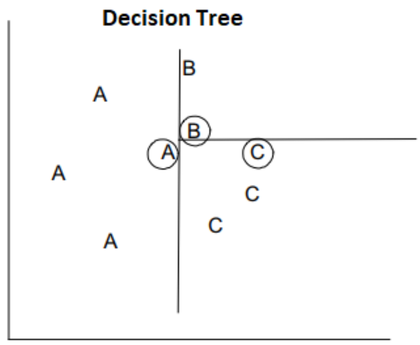
Question 6. [Marks 1+1+2=4] **[to be answered only on pages 15-16]**

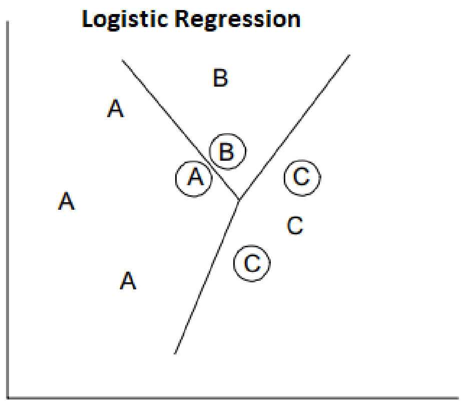
Draw the decision boundary (shape and position w.r.t. training points labelled as class A, B, and C in the figure below) for Decision Tree, Logistic Regression and Gaussian Naïve Bayes (different means and different variances for different classes) classifiers.

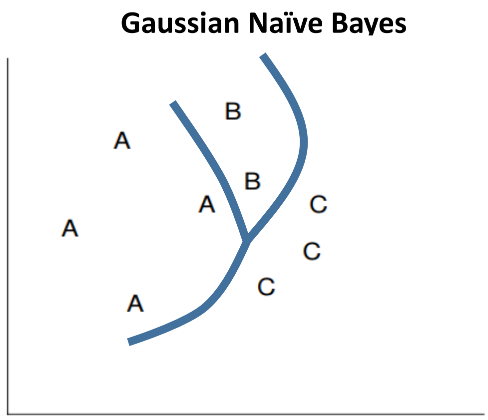
Decision boundary for Decision tree and Logistic Regression: piecewise linear (straight line) between A & B, A&C and B&C.

For decision trees, the boundary is axis-aligned, either horizontal or vertical.

Decision boundary for Gaussian Naïve Bayes (different variance for different classes as per the given training data): piecewise quadratic (non-linear curve) between A & B, between A & C, and between B & C.


Decision Tree
B
A
B
A
C
A
C
C
A


Logistic Regression
B
A
B
A
C
A
C
C
A


Gaussian Naïve Bayes
B
A
B
A
C
A
C
C
A