CS3244 Machine Learning – Final (New Topics) Sample Answers

Section 1: Model Evaluation

Q1-2: A car repair company is using an audio AI system to diagnose whether car engines are faulty or normal (Prediction). For comparison, a senior mechanic also examined the engines and rated them (Actual). These are the results from 10 cars.

Car ID	Prediction	Actual
1	Normal	Normal
2	Normal	Normal
3	Normal	Normal
4	Faulty	Normal
5	Normal	Normal
6	Faulty	Normal
7	Normal	Normal
8	Faulty	Faulty
9	Normal	Faulty
10	Faulty	Faulty

Q1) [4 Marks] MRQ: What metric(s), which is not misleading, should you use to report how well the model is doing?

a) Cosine Distance
b) Accuracy
c) Recall of Faulty prediction
d) Precision of Faulty prediction

Answer: c

Justification / Working:

a) False. Cosine similarity is measured for vector regression.
b) False. Since the data is imbalanced, accuracy is misleading.
c) True. Car engine diagnosis is similar to rare disease diagnosis. Missing to diagnose problems may cause accidents, hence false negative is very costly. Recall $R = TP/(TP + FN)$ is more important to decrease.
d) False. Prediction $P = TP/(TP + FP)$. But False Positive (FP) is less important, since this just means some wasted work.

Q2) [4 Marks] Calculation: Calculate the F_1 score for the model.

4/7 or 0.5714

Justification / Working:

Positive = Faulty, Negative = Normal.

$$TP = 2, TN = 5, FP = 2, FN = 1$$

$$P = \frac{TP}{TP + FP} = \frac{2}{2 + 2} = \frac{1}{2}$$

$$R = \frac{TP}{TP + FN} = \frac{2}{2 + 1} = \frac{2}{3}$$

$$F_1 = \frac{2}{\frac{1}{P} + \frac{1}{R}} = \frac{2}{2 + \frac{3}{2}} = \frac{4}{7} = 0.5714$$

Section 2: Data Processing and Feature Engineering

Q3) [2 Marks] T/F: PCA provides better features for regression than LDA.

a) True
b) False

Answer: **False**

Justification / Working:

LDA is trained with labels, so it helps to learn good features for supervised learning tasks (classification and regression).
PCA is not trained with labels, so it may not help towards a specific prediction task.

Q4) [6 Marks] Calculation: Using Tokenization, Stop Word Filtering, Lemmatization, and Bag-of-Words, calculate the distance between the numeric representations of the two sentences $x^{(1)}$ and $x^{(2)}$.

Specify the distance you used and the number in this format:
Simple Difference = 123

Stop words: ["a", "any", "as", "by", "just", "other", "that", "we", "which"]

$x^{(1)}$: "A rose by any other name smells just as sweet."
$x^{(2)}$: "That which we call a rose by any other word would smell as sweet."

Absolute difference = 4
Euclidean distance = 2
Cosine similarity = 1/8, Angular distance = 1.45

Justification / Working:

Filtered Stop Words:
$x^{(1)}$: "~~A~~ rose ~~by~~ ~~any~~ ~~other~~ name smells ~~just~~ ~~as~~ sweet."
= ["rose", "name", "smells", "sweet"]
$x^{(2)}$: "~~That~~ ~~which~~ ~~we~~ call ~~a~~ rose ~~by~~ ~~any~~ ~~other~~ word would smell ~~as~~ sweet."
= ["call", "rose", "word", "would", "smell", "sweet"]

Lemmatization: "smells" → "smell"

Bag-Of-Words:

w	rose	name	smell	sweet	call	word	would
$\boldsymbol{x}_w^{(1)}$	1	1	1	1	0	0	0
$\boldsymbol{x}_w^{(2)}$	1	0	1	1	1	1	1
$\left\lvert \boldsymbol{x}_w^{(1)} - \boldsymbol{x}_w^{(2)} \right\rvert$	0	1	0	0	1	1	1

w	rose	name	smell	sweet	call	word	would
$\boldsymbol{x}_w^{(1)}$	1	1	1	1	0	0	0
$\boldsymbol{x}_w^{(2)}$	1	0	1	1	1	1	1
$\boldsymbol{x}_w^{(1)}\boldsymbol{x}_w^{(2)}$	1	0	1	1	0	0	0

Distance if using absolute difference (Manhattan Distance):

$$d\left(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}\right) = \left\|\boldsymbol{x}_w^{(1)} - \boldsymbol{x}_w^{(2)}\right\|_1 = \sum_w \left|\boldsymbol{x}_w^{(1)} - \boldsymbol{x}_w^{(2)}\right| = 4$$

Distance if using Euclidean difference:

$$d\left(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}\right) = \left\|\boldsymbol{x}_w^{(1)} - \boldsymbol{x}_w^{(2)}\right\|_2 = \sqrt{\sum_w \left(\boldsymbol{x}_w^{(1)} - \boldsymbol{x}_w^{(2)}\right)^2} = 2$$

Cosine Similarity:

$$s = \cos\theta = \frac{\boldsymbol{x}_w^{(1)} \cdot \boldsymbol{x}_w^{(2)}}{\left\|\boldsymbol{x}_w^{(1)}\right\|_2 \left\|\boldsymbol{x}_w^{(1)}\right\|_2} = \frac{3}{4 \times 6} = \frac{1}{8}$$

Angular Distance (calculate Cosine Similarity first):

$$d\left(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}\right) = \cos^{-1} s = 1.45$$

Justification / Working (without lemmatization):

Filtered Stop Words:

$x^{(1)}$: "~~A~~ rose ~~by any other~~ name smells ~~just as~~ sweet."
= ["rose", "name", "smells", "sweet"]

$x^{(2)}$: "~~That which we~~ call ~~a~~ rose ~~by any other~~ word would smell ~~as~~ sweet."
= ["call", "rose", "word", "would", "smell", "sweet"]

Bag-Of-Words:

w	rose	name	smells	sweet	call	word	would	smell
$x_w^{(1)}$	1	1	1	1				
$x_w^{(2)}$	1		1	1	1	1	1	1
$\left\|x_w^{(1)} - x_w^{(2)}\right\|$		1			1	1	1	1

Section 3: Explainable AI

Q5) [2 Marks] T/F: Explaining a logistic regression model using LIME with a sparse linear model will produce identical weights.

a) True
b) False

Answer: **False**

Justification / Working:

The logistic regression model is globally trained on the whole dataset, but LIME trains the explanation model only for a local neighborhood. So, their model weights are unlikely to be the same.

--- End ---