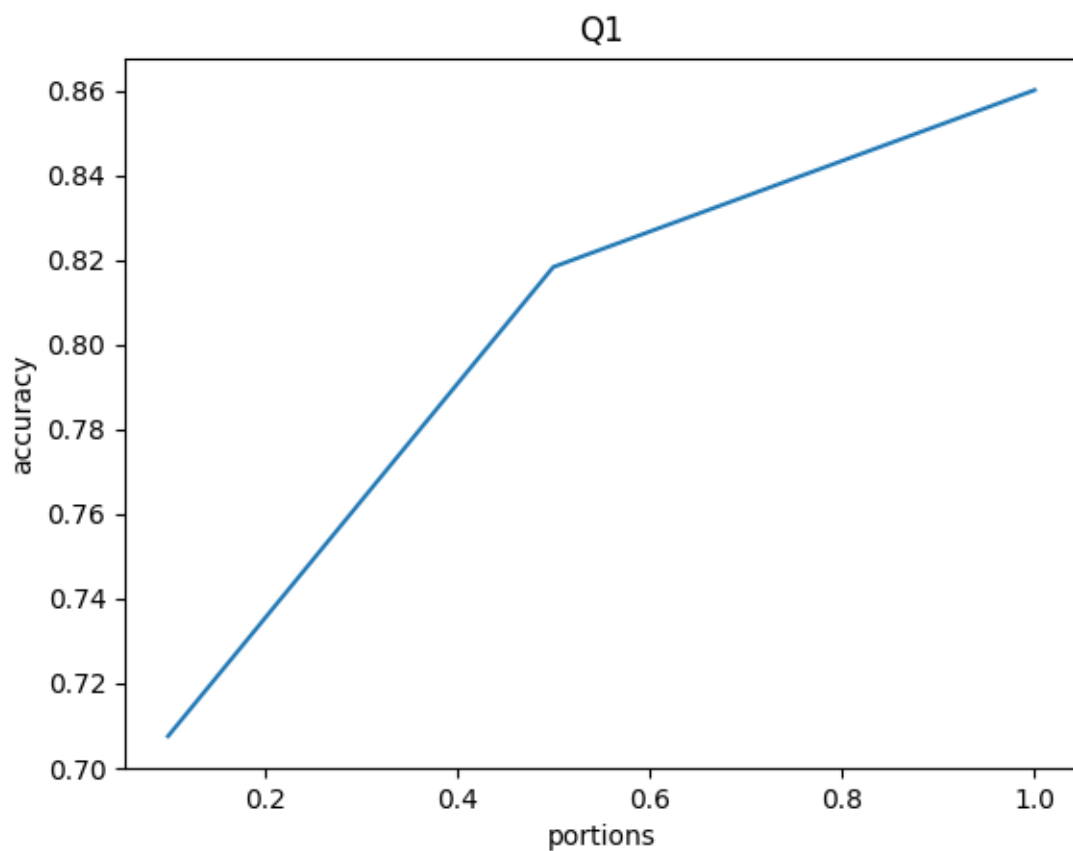


תרגיל 5 | NLP (67658)

שם: רונאל חרדים, עומרי טויטו | ת"ז: 208472761, 208917641

שאלה 1 - Linear Regression:

גרף של accuracy כפונקציה של האחוז מהדאטה:



התוצאות שהתקבלו הן:

Logistic regression results:

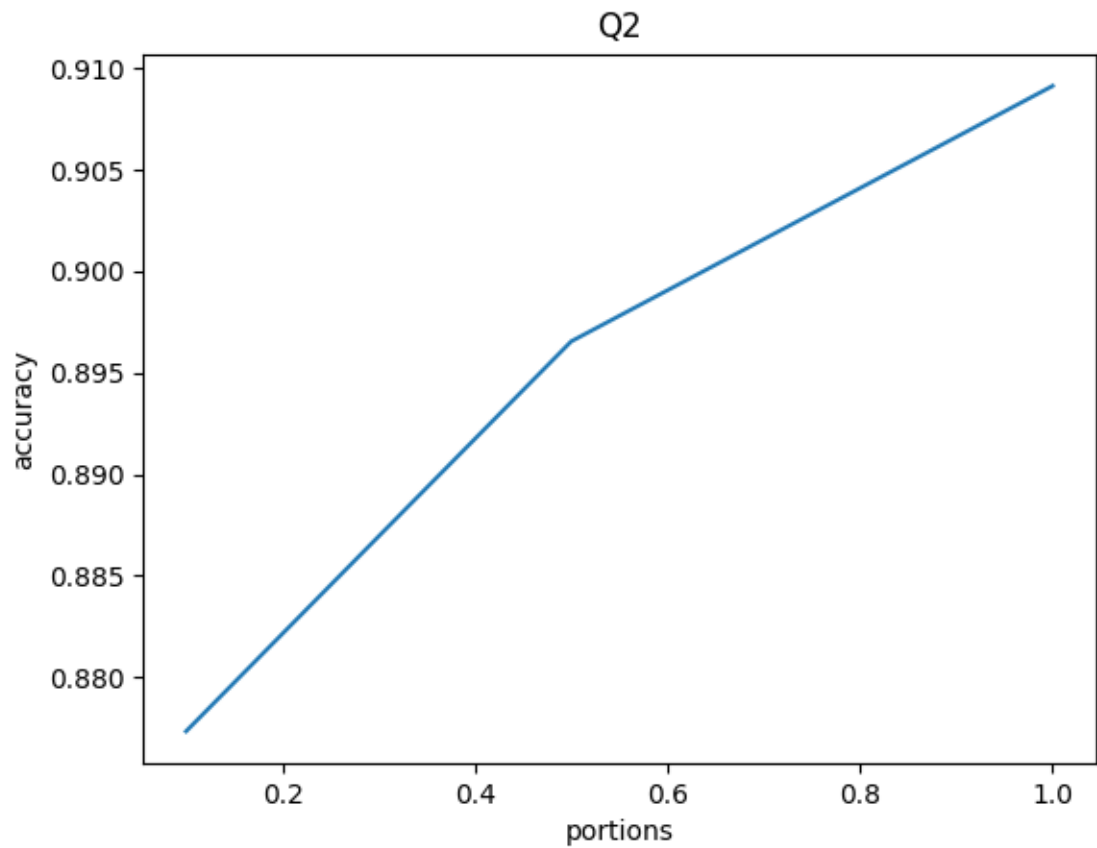
Portion: 0.1 - 0.7075596816976127

Portion: 0.5 - 0.8183023872679045

Portion: 1.0 - 0.860079575596817

שאלה 2 - SequenceClassification:

גרף של accuracy כפונקציה של האחוז מהדאטה:



התוצאות שהתקבלו הן:

Finetuning results:

Portion: 0.1: 0.8773209549071618

Portion: 0.5: 0.896551724137931

Portion: 1.0: 0.9091511936339522

שאלה 3 - zero-shot:

ה *accuracy* שקיבלנו היא:

Zero – shot result : 0.7712201591511937

שאלה 4

(א)

המודל שקיבל את ה *accuracy* הגבוה ביותר הוא: המודל השני SequenceClassification

(ב)

המודל שהיה הכי רגיש לגודל הדאטה סט הוא: המודל הראשון Linear Regression

(ג)

נביא שני יתרונות וחסרונות של מודל zero-shot:

יתרונות:

1. מודל זה יכול לסווג מחלקות חדשות מבלי שהא ראה את התיוג שלהן לפניכן. זה יכול לחסוך לנו זמן ומשאבים של תיוג.

2. הוא יכול להכליל טוב יותר, משום שהוא משתמש בידע מוקדם על הקשרים בין המחלקות.

חסרונות:

1. הוא צריך לקבל כמות ענקית של מידע כדי להכליל טוב, ולהבין את ההבדלים בין המחלקות השונות.

2. הוא סובל מאי וודאות, אי אפשר לדעת עבור אלו משימות הוא טוב עד שננסה ונבדוק. לעומת מודלים האחרים שניתן להציע להם *fine tuning* לכמה מחלקות שונות.