

Deep Learning

Amit Waizman & Rivka Buskila

אטרא

לסוג חדש לקטגוריות שונות על סמך התוכן שלהם

Dataset

0: Politics

1: Sports

2: Technology

3: Entertainment

4: Business

Label	Number of Examples
0	417
1	511
2	401
3	386
4	510

Feature Extraction

עיבוד טקסט לערכים מספריים לשילוג באמצעות Bag of Words להמרת טקסט לוקטור מספרי.

שלב ראשון: ייצרת מילון (Vocabulary) מכל הטקסטים.

שלב שני: ספירת הופעות של מילים מהמילון בכל מסמך.

שלב שלישי: עיבוד טקסטים באמצעות NLTK:

הסרת מילים נפוצות.

הסרת סימני פיסוק.

המרת מילים לצורה בסיסית.

גודל אוצר המילים (Vocabulary Size): 24,331 מילים.

The background features a dark purple gradient with three semi-transparent, rounded rectangular shapes. One large shape is positioned on the left side, another smaller one is at the top center, and a third medium-sized one is at the bottom center.

Baseline

Baseline

יצירת מודל פשוט שמסוווג את כל הנתונים לקטgorיה הנפוצה ביותר.

תוצאות:

Metric	Test Set (%)	Validation Set (%)
Accuracy	20.67	24.49
Precision	4.27	6.00
Recall	20.67	24.49



SoftMax

SoftMax

תהליכי האימון:

.0.001 = (Learning Rate)
.epochs 2,000

תוצאות:

Metric	Test Set (%)	Validation Set (%)
Accuracy	97.30	93.93
Precision	97.35	94.01
Recall	97.30	93.93



SoftMax

תהליכי האימון:

שיעור למידה (Learning Rate) .01

.epochs 1,000

Ridge Regularization

תוצאות

Metric	Test Set (%)	Validation Set (%)
Accuracy	96.63	95.28
Precision	96.73	95.31
Recall	96.63	95.28

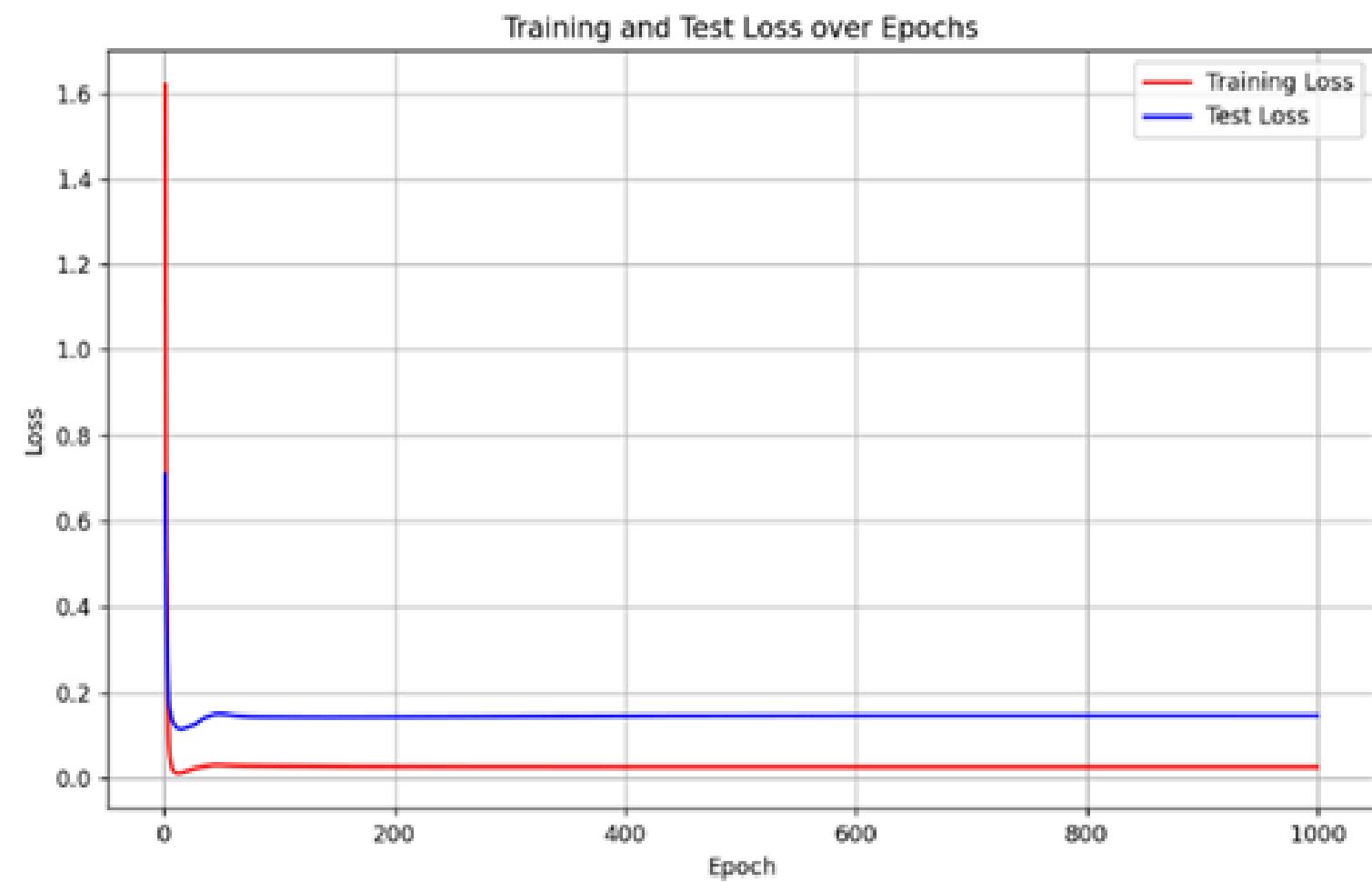
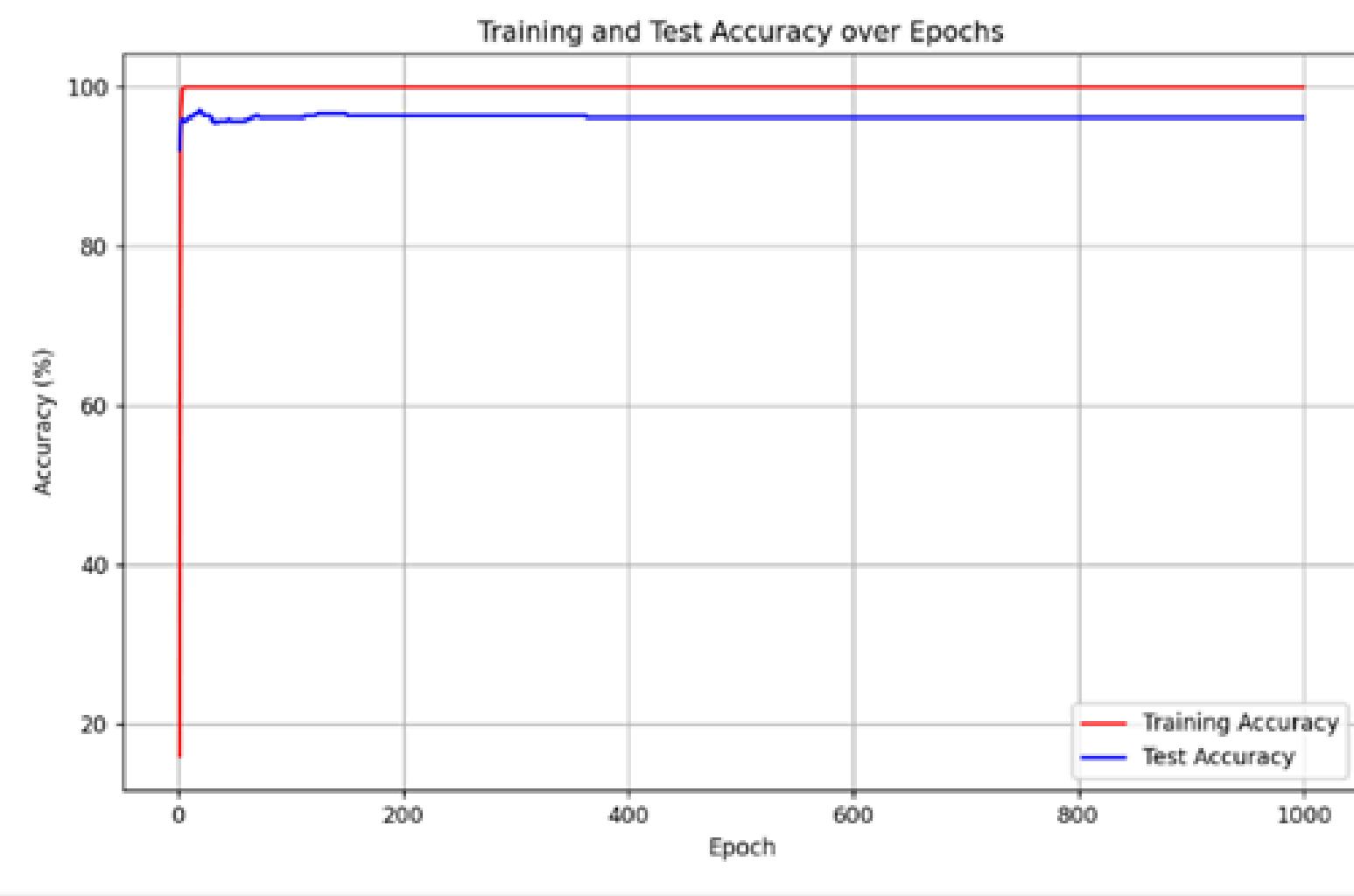


SoftMax

תהליכי האימון:

שיעור למידה (Learning Rate) .0.01 =
התכנסות אחרי epochs 10.

Ridge Regularization
במקומ SGD השתמשנו Adam Optimizer



SoftMax

תהליך האימון:

שיעור למידה (Learning Rate) = 0.01
התכנסות אחרי 10 epochs.

Ridge Regularization

במקום SGD השתמשנו Adam Optimizer

תוצאות

Metric	Test Set (%)	Validation Set (%)
Accuracy	97.08	96.18
Precision	97.17	96.18
Recall	97.08	96.18



Neural Network

Neural Network

3 שכבות:

- **שכבה קלט**: בגודל אוצר המילים (24,331).
- **שכבה חבויה**: בגודל 12 (נבחר בניסוי וטעייה).
- **שכבה פלט**: בגודל כמות הקטגוריות (5).

תוצאות עבור 1000 איטרציות

Metric	Test Set (%)	Validation Set (%)
Accuracy	97.3	95.28
Precision	97.35	95.3
Recall	97.3	95.28



Neural Network

הוספה שכבה נוספת:

- **שכבה קלט**: בגודל אוצר המילים (24,331).
- **שכבה חבוייה 1**: בגודל 20.
- **שכבה חבוייה 2**: בגודל 10.
- **שכבה פלט**: בגודל כמות הקטגוריות (5).

תוצאות עבור 1000 איטרציות

Metric	Test Set (%)	Validation Set (%)
Accuracy	97.3	96.18
Precision	97.36	96.17
Recall	97.3	96.18



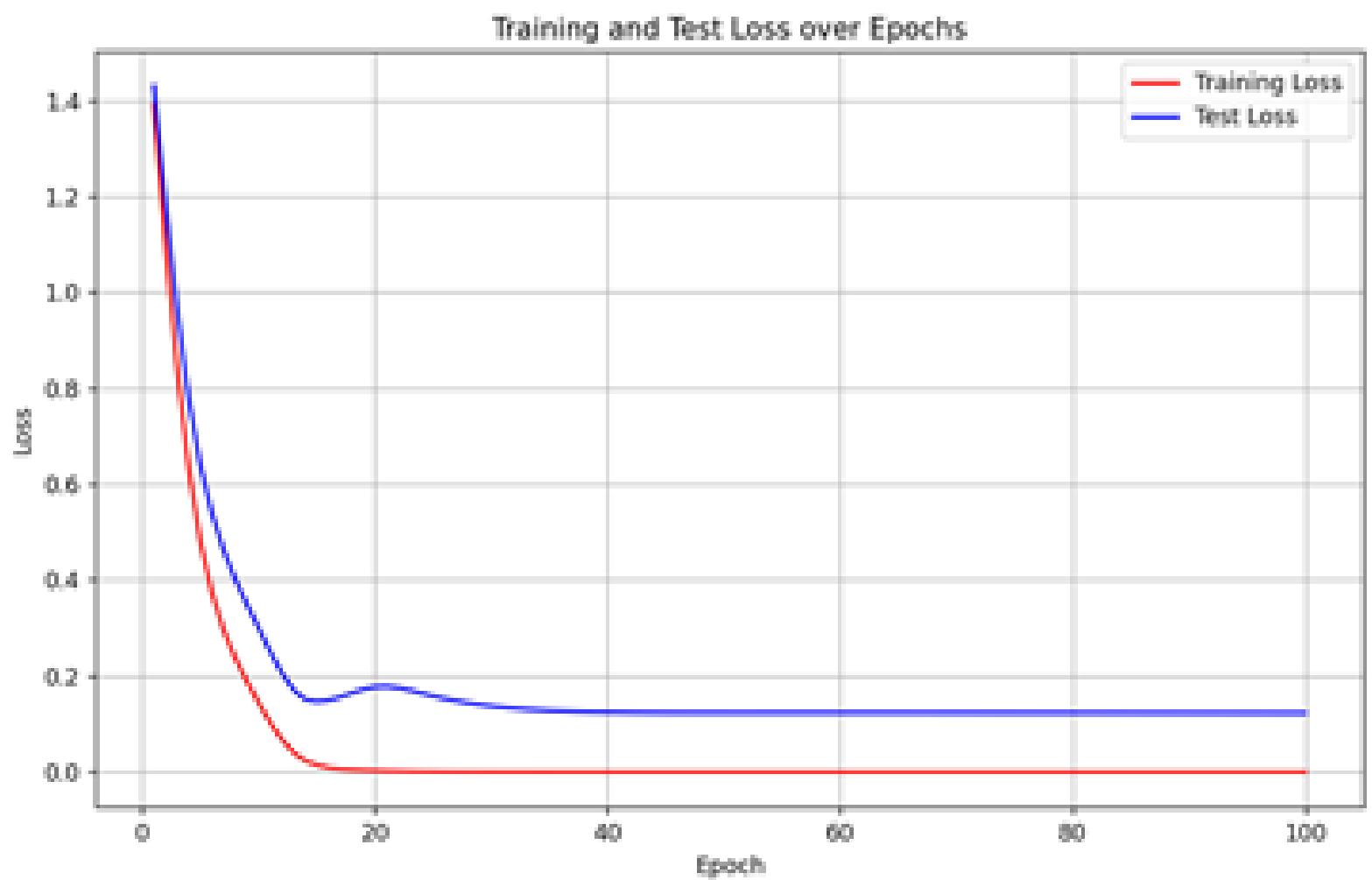
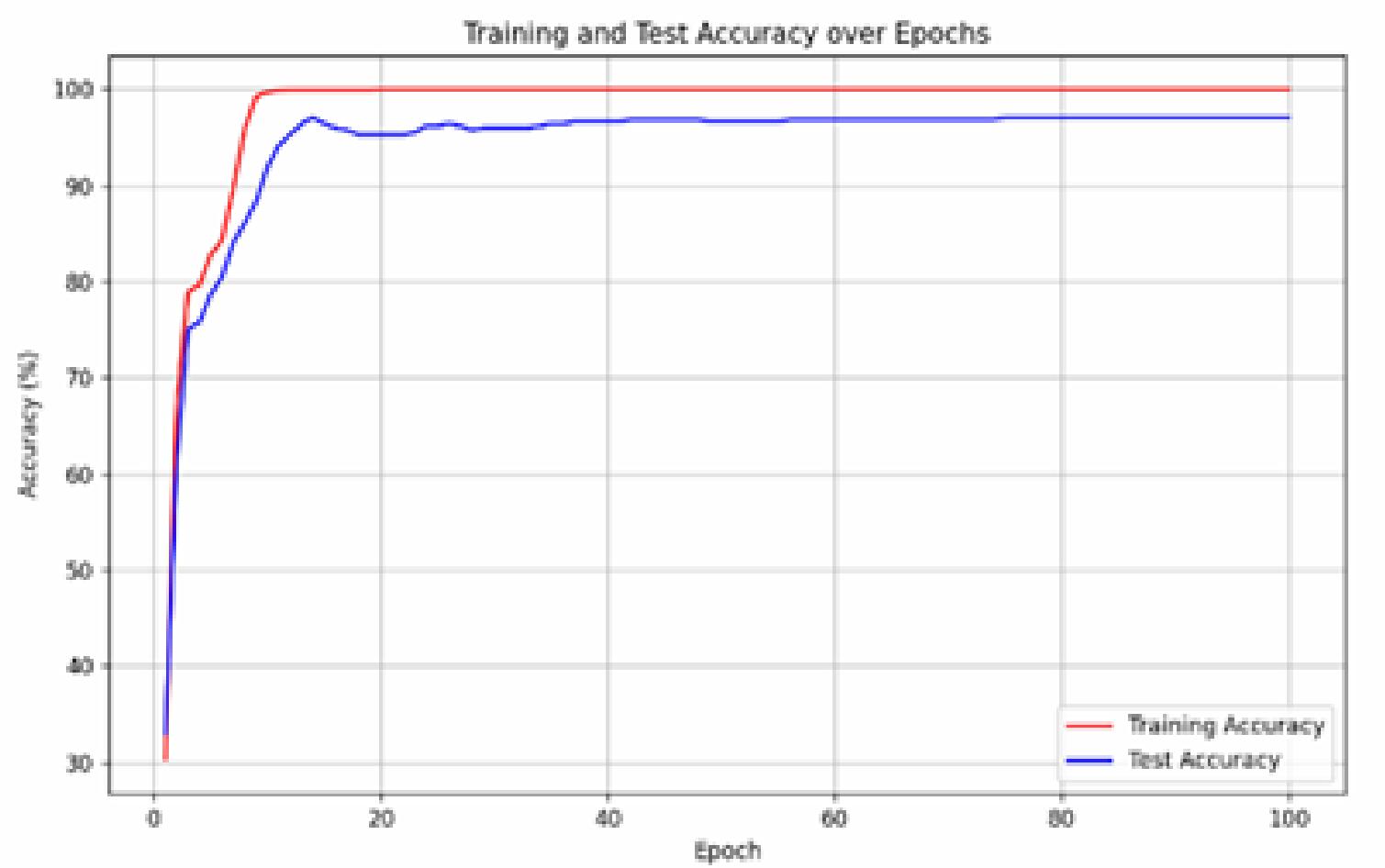
Compare Neural Network to SoftMax

SGD by

Metric	Test Set (Neural Network)	Validation Set (Neural Network)	Test Set (SoftMax)	Validation Set (SoftMax)
Accuracy	97.3	96.18	96.63	95.28
Precision	97.36	96.17	96.73	95.31
Recall	97.3	96.18	96.63	95.28

עכירה לאחר epoch 15.

שימוש ב-Adam Optimizer



Neural Network

שימוש ב-Adam Optimizer

4 שכבות

Metric	Test Set (%) Neural Network	Validation Set (%) Neural Network
Accuracy	97.75	97.98
Precision	97.79	98.02
Recall	97.75	97.98



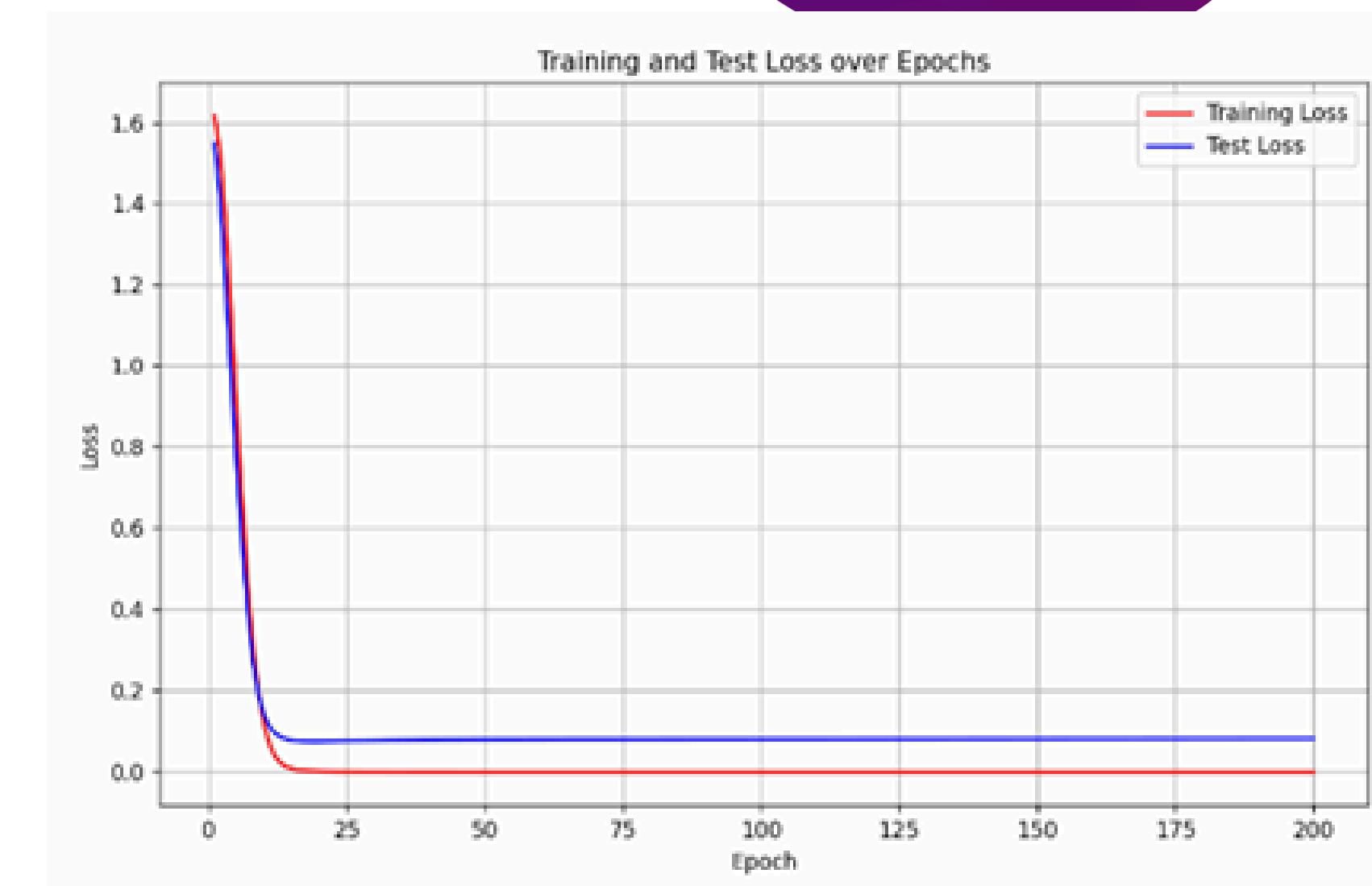
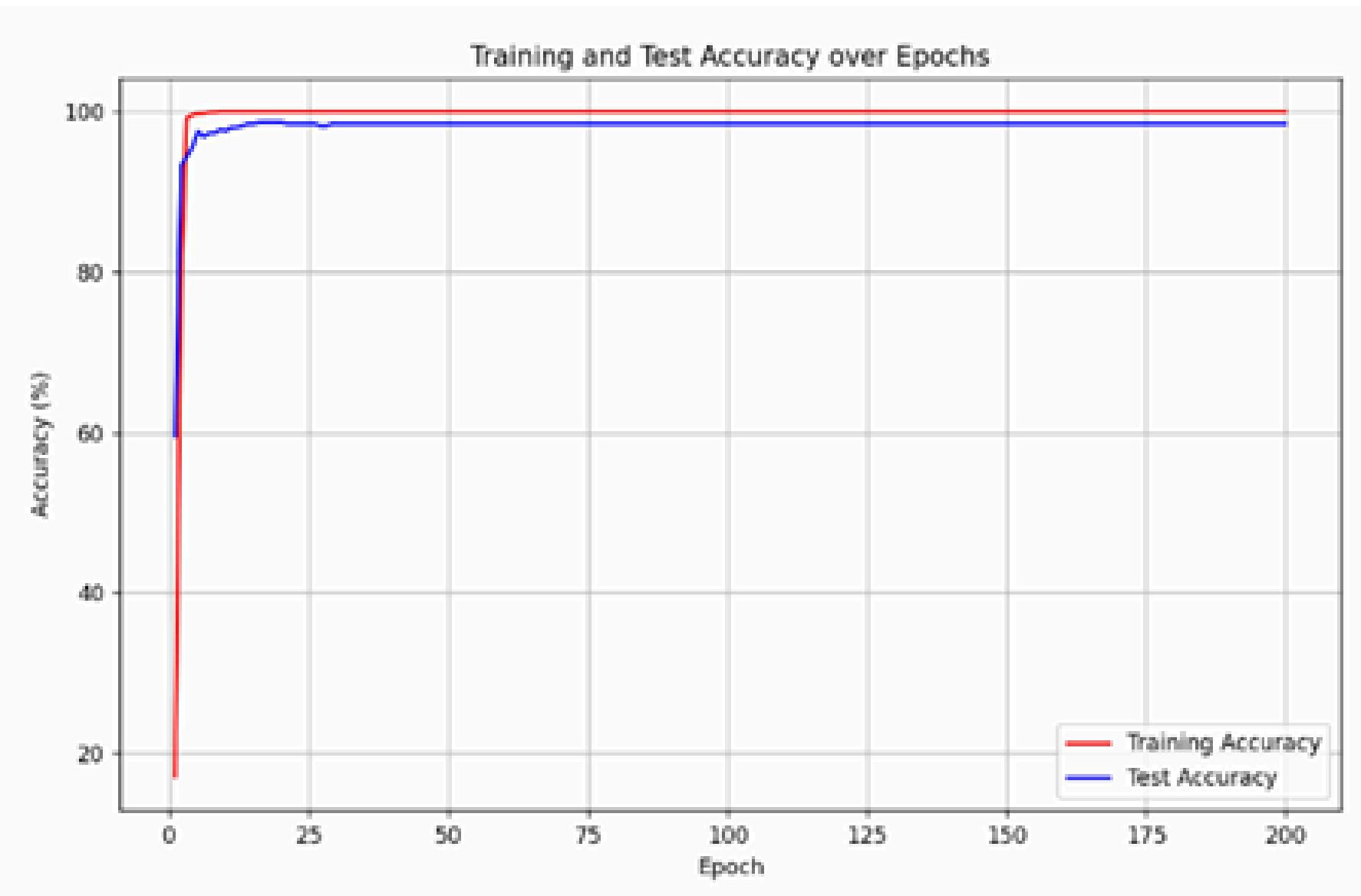
Compare Neural Network to SoftMax

Adam DY

Metric	Test Set (%) Neural Network	Validation Set (%) Neural Network	Test Set (%) SoftMax	Validation Set (%) SoftMax
Accuracy	97.75	97.98	97.08	96.18
Precision	97.79	98.02	97.17	96.18
Recall	97.75	97.98	97.08	96

LSTM

Experiment	Epochs	Neurons	Layers	Optimizer	Test Set (Accuracy)	Validation Set (Accuracy)
First Experiment (Word2Vec)	1000	64	1	Adam	94.61%	92.36%
First Experiment (Word2Vec)	200	64	1	Adam	92.13%	91.69%
Second Experiment (Bag of Words)	100	64	1	Adam	97.75%	98.2%
Second Experiment (Bag of Words)	200	64	2	Adam	98.43%	97.98%
Second Experiment (Bag of Words)	20	64	2	Adam	97.75%	98.43%



תודה על ההקשבה