

## חלק ב

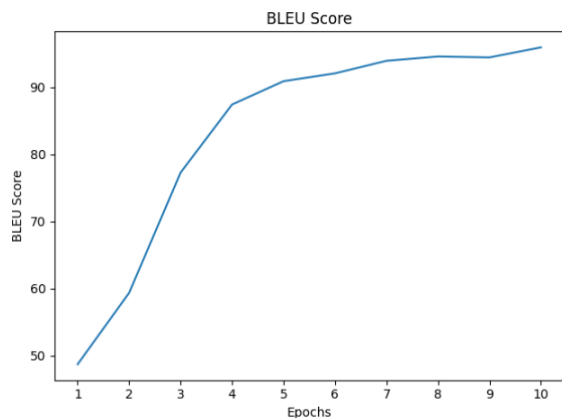
המקודד מורכב משכבת STM חד כיוונית יחידה עם מימד של 128 שכבות נסתרות.  
מימד האמבדינג הוא 128.  
אופטימיזציה ADAM.  
NLL Loss.  
Learning rate של 0.00020

המפענח מורכב משכבת STM חד כיוונית יחידה עם מימד של 256 שכבות נסתרות.  
מימד האמבדינג הוא 128.  
אופטימיזציה ADAM.  
NLL Loss.  
Learning rate של 0.00020  
הפלט של ה-LSTM עובר דרך שכבת dropout עם rate של 0.3

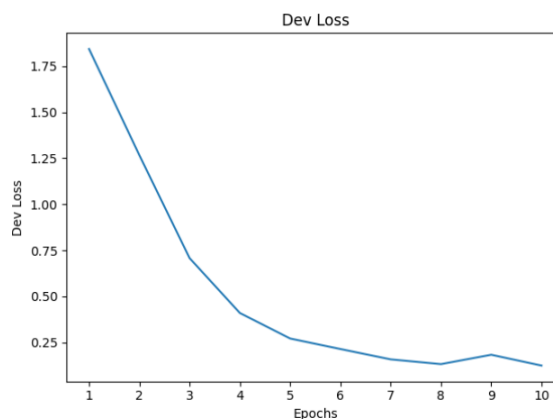
זמן האימון: 10.865 דקות  
ניתן לראות כי זמן האימון במודל זה כפול מהזמן של המודל הבסיסי שמימשתי בחלק 1 של התרגיל.  
הסיבה לכך היא שבמודל זה אנחנו מחשבים את ה-attention.

Train Loss = 0.051

BLEU Score = 95.930



Dev Loss = 0.124



כאשר הרצתי את evaluation.py

עם הפרמטרים של test קיבלתי ש-BLEU Score = 96.772

כלומר, ניתן לראות שיפור משמעותי מהתוצאה של המודל הבסיסי בו קיבלתי את התוצאה:  
BLEU Score = 80.627

לסיכום,

הביצועים של מודל זה טובים יותר מהמודל הבסיסי מסעיף 1.  
הערכים שהתקבלו במודל זה טובים משמעותית מהערכים שהתקבלו במודל הבסיסי, גם ערך ה-Dev Loss וגם BLEU Score בשלב האימון ובשלב הטסט.  
לכן Attention-Based Decoder Model טוב יותר מ-Basic Encoder-Decoder Model.