

Jacob Haywood

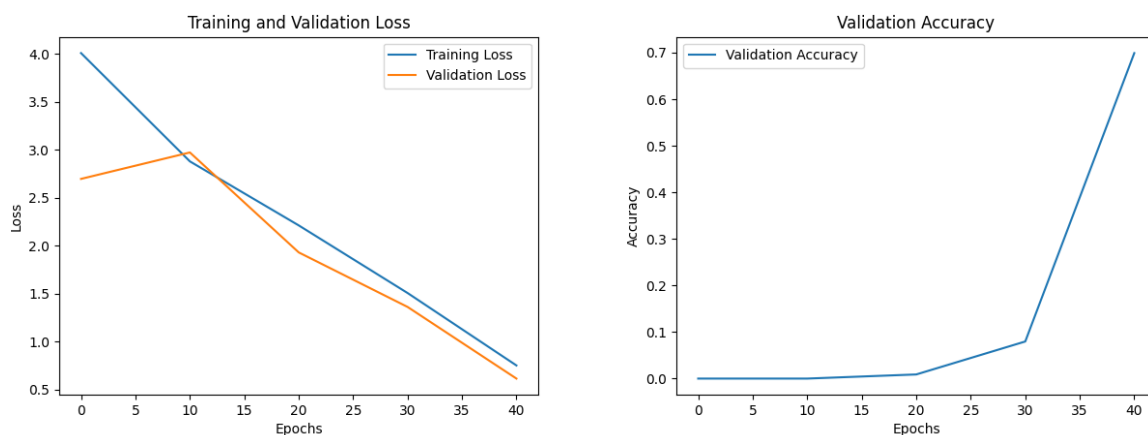
801182240

Assignment 4

Github: https://github.com/jhaywoo6/Deep_Learning

The given assignment was to design, test, and evaluate a basic GRU-based encoder-decoder for translating common English phrases to French and vice versa. The objective being to determine the difficulty translating such languages and determining the effect of attention on the final validation loss and accuracy. While an option for GRU's from pytorch, bi-directional models will not be tested here. During experimentation it was found that with the given dataset it was possible to create all four models in such a way that they would have nearly 100% accuracy and have negligible differences, so parameters have been selected to simplify the models and emphasize the effect. Specifically, a hidden size of 128, batch size 1, and 1 layer over 41 epochs were selected for all four models. Models 1 and 2 translate text from English to French and compare the effect of no attention with attention, with models 3 and 4 studying the inverse. Both models use the same encoder, and the decoders for attention and non-attention were built to be as similar as possible apart from the addition of attention. Word based prediction is used here as opposed to character-based prediction.

Problem 1: Model , English to French without attention



Input: They visit museums often, Target: Ils visitent souvent des musées, Predicted: Ils visitent souvent des musées

Input: We watch a movie together, Target: Nous regardons un film ensemble, Predicted: Nous regardons un film ensemble

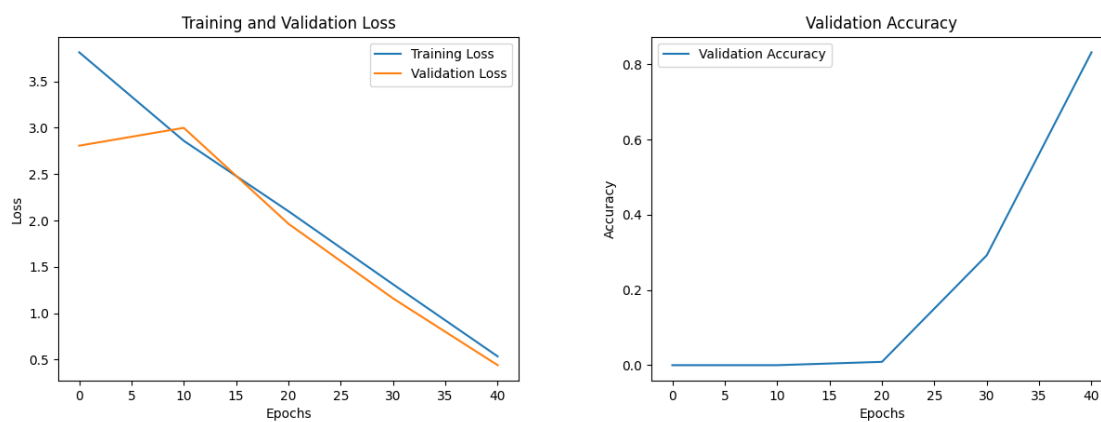
Input: They play video games, Target: Ils jouent aux jeux vidéo, Predicted: Ils jouent aux jeux vidéo

Input: She studies hard for exams, Target: Elle étudie dur pour les examens, Predicted: Elle étudie dur pour les

Input: She ties her shoelaces, Target: Elle attache ses lacets, Predicted: Elle attache ses lacets

Final Validation Loss and Accuracy: 0.613 and 69.9%

Problem 2: Model 2, English to French with attention



Input: She swims in the ocean, Target: Elle nage dans l'océan, Predicted: Elle nage dans l'océan

Input: He writes a letter, Target: Il écrit une lettre, Predicted: Il écrit une lettre

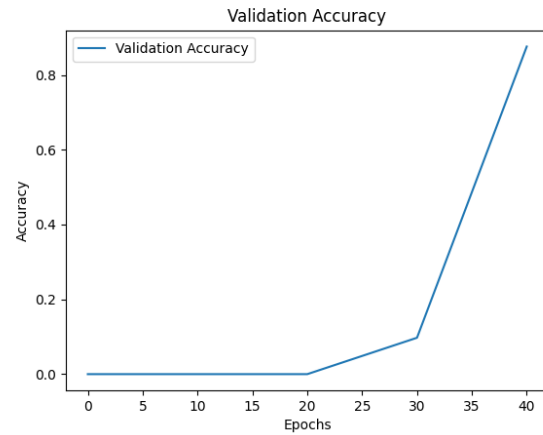
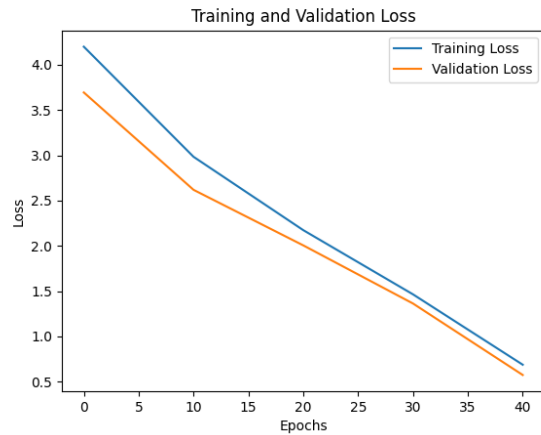
Input: We watch movies on Fridays, Target: Nous regardons des films le vendredi, Predicted: Nous regardons des films le vendredi

Input: They go to the gym, Target: Ils vont à la salle de sport, Predicted: Ils vont à la salle de la

Input: We visit the museum, Target: Nous visitons le musée, Predicted: Nous visitons le musée
Final Validation Loss and Accuracy: 0.44 and 83%

Applying attention to the model appears to have made training more efficient within the same model parameters compared to the model without attention, experiencing a sharper drop off in loss and acquiring accuracy at a faster rate.

Problem 3: Models 3 & 4, French to English



Input: Le chien aboie bruyamment, Target: The dog barks loudly, Predicted: The dog barks loudly

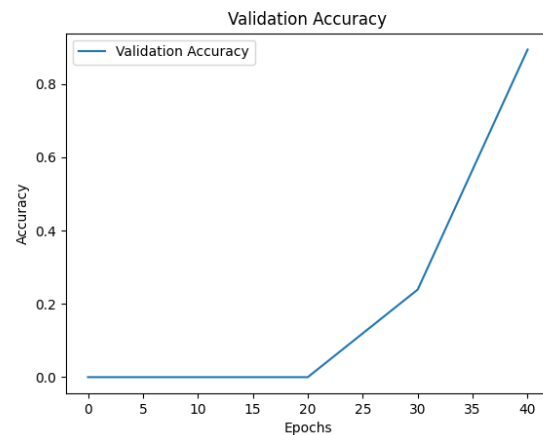
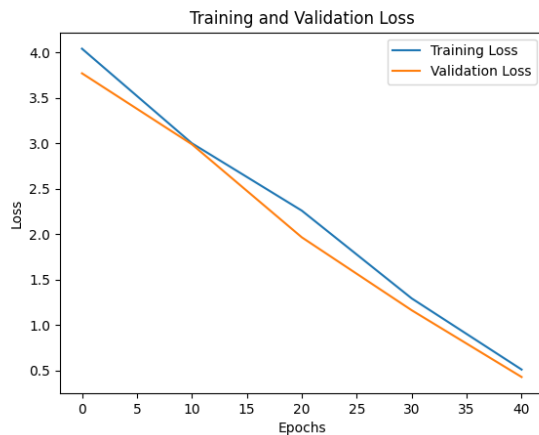
Input: Ils apprécient le coucher du soleil, Target: They enjoy the sunset, Predicted: They enjoy the sunset

Input: Ils font du vélo, Target: They ride bicycles, Predicted: They ride bicycles

Input: Nous construisons un château de sable, Target: We build a sandcastle, Predicted: We build a sandcastle

Input: Il se lave les mains, Target: He washes his hands, Predicted: He washes his hands

Final Validation Loss and Accuracy: 0.573 and 87%



Input: Le vent souffle doucement, Target: The wind blows gently, Predicted: The wind blows gently

Input: Elle chante une chanson, Target: She sings a song, Predicted: She sings a song

Input: Il éteint la lumière, Target: He turns off the light, Predicted: He turns off the light

Input: Nous prenons le petit déjeuner ensemble, Target: We eat breakfast together,
Predicted: We cook breakfast together

Input: Il écrit une lettre, Target: He writes a letter, Predicted: He writes a letter

Final Validation Loss and Accuracy: 0.428 and 89%

Translating in reverse appears to be a much simpler process, resulting in a higher and tighter accuracy and loss range for both non-attention and attention models. There is also no initial loss jump after 10 epochs compared to models 1 and 2 which is interesting to note. The reasoning for this is likely extra complexity in French sentences compared to English counterparts, with English sentences being simpler to generate compared to French ones, requiring a mildly more complex model to achieve the same results as models 3 and 4 did here.