

코드는 Google Colab 환경을 기준으로 작성되었다.

Glove data load:

아래의 `read_glove_vecs()`에 해당 glove 파일이 있는 경로로 수정하면 불러올 수 있다.

```
words_to_index, index_to_words, word_to_vec_map = read_glove_vecs('/content/mnt/My Drive/ugrp/glove.6B.50d.txt')
X, Y = read_csv()
label_to_emoji(Y[0])
```

`LSTM_ADAM_Model`, `LSTM_SGD_Model`, `LSTM_SGD_Drop_Model`, RNN 총 4 가지의 큰 모델이 class 로 구성되어있으며, 아래의 train 코드에 해당 객체의 파라미터로 glove의 dimension 을 넘겨주면 구체적인 모델을 불러올 수 있다.

Ex)

```
LSTM+SGD+50d: model = LSTM_SGD_Model(lr,emb_size=50)

LSTM+ADAM+50d: model = LSTM_ADAM_Model(lr,emb_size=50)

LSTM+SGD+100d: model = LSTM_SGD_Model(lr,emb_size=100)

lSTM+SGD+50d+dropout: model = LSTM_SGD_Drop_Model(lr,emb_size=50)

RNN+SGD+50d: model = RNN(len(word_to_vec_map), 50, 128)
```

RNN의 경우, 이미 RNN+SGD+50d의 조건으로 구현된 모델이므로 따로 수정하지 않고 사용할 수 있다.

```
%%time

mini_batch = 2
lr = 0.01

total_loss = 0
loss_count = 0
train_loss = []
valid_loss = []
v_loss = 0

model = LSTM_SGD_Drop_Model(lr=0.7, emb_size = 50) #50dim
idx = (len(X_train) - 1) // mini_batch #131/6 = 21 #21개의 batch 묶음을 만들겠다!
offsets = [i * idx for i in range(mini_batch)]
v_x = np.concatenate((X_train_valid, X_train[idx:len(X_train_valid)]), axis=0)
v_y = np.concatenate((Y_train_valid, Y_train[idx:len(X_train_valid)]), axis=0)
print('validation_x: ', v_x.shape, 'validation_y: ', v_y.shape)
for epoch in range(20): #epoch 수
    print('epoch: ', epoch)
    total_loss = 0
```

## 1. Accuracy comparison for test set

lr = learning rate

(1) RNN+SGD+50d epoch = 1000 lr = 0.4	33.92857142857143%
(2) LSTM+SGD+50d epoch: 1000 lr = 0.5	71.42857142857143%

(3) LSTM+ADAM+50d epoch: 20 lr = 0.01	64.28571428571429
(4) LSTM+SGD+100d epoch: 1000 lr = 0.5	50%
(5) LSTM+SGD+50d+dropout epoch: 1000 lr = 0.5	66.07142857142857%

동일한 모델이었을 때 SGD의 accuracy가 ADAM보다 높은 것을 확인할 수 있다. 다만, 이는 SGD의 learning rate를 0.7이라는 높은 숫자로 설정하고, Adam은 그에 반해 0.01이라는 아주 작은 숫자로 설정한 차이와 SGD의 epoch은 1000이지만 adam은 20epoch으로도 그와 준하는 accuracy를 낸 차이가 있다. 즉, SGD보다 ADAM이 적은 에폭과 적은 learning rate로도 좋은 accuracy를 도출해냈다. 이는 ADAM의 특성과 관련있는데, ADAM은 RMSProp과 Momentum 방식을 합친 것과 같은 알고리즘이다. 그렇기에 Adam의 파라미터 갱신 과정은 그릇 바닥을 구르듯이 움직이며, 결과적으로 모멘텀보다 움직임의 좌우 흔들림이 적다. 그렇기에 이번 모델에서 Adam을 사용하면 더 적은 수의 에폭으로 좋은 성능을 낼 수 있다.

RNN의 경우, accuracy가 가장 낮게 나왔다. 10,100,500 epoch에서도 20~30%의 accuracy가 나왔기에 1000 epoch까지 돌려 모델이 과적합될 위험성을 감수하더라도 test set에 대한 accuracy를 높여보았다. 그래서 1000 epoch 이후 추가적으로 2000 epoch을 더 학습하여 총 3000 epoch에서의 성능도 확인해보았지만 오히려 모델이 train set에 더 과적합되어서 accuracy가 감소하였다. Learning rate를 줄이거나 높여도 accuracy 상의 유의미한 차이를 얻을 수 없었다. 따라서, 이는 기본적인 모델의 한계이거나 학습 데이터 양의 부재일 수도 있다.

Train data는 106개, validation data는 26개, test data는 56개로 train data의 수가 상당히 적은 것을 알 수 있다. 이러한 train data가 test data를 잘 예측할만큼 다양성을 확보한 train data라고 볼 수 없으며, epoch을 높일수록 모델은 점점 더 train data에 과적합될 수밖에 없다.

또한, 생각보다 Dropout의 성능을 확인할만한 Accuracy는 나오지 않았다. 오히려 Dropout을 사용하지 않은 모델이 더 좋은 성능을 보였다. 이 역시 기본적인 학습 데이터 양으로 인한 모델의 한계 때문일 수 있다.

## 2. All emojis for test set

위의 1번 문항의 5개의 모델들의 test set emoji의 출력 결과물이다.

















































































































양식: [test문]      [모델이 예측한 emoji]    [실제 emoji]

(1) RNN+SGD+50d	<p>             I want to eat 🤔👤              he did not answer 😞😞              he got a very nice raise 🤔😞              she got me a nice present 😞😞              ha ha ha it was so funny 😞😞              he is a good friend 😞😞              I am upset 😞😞              We had such a lovely dinner tonight 😞😞              where is the food 🤔👤              Stop making this joke ha ha ha 😞😞              where is the ball 🤔👤              work is hard 😞😞              This girl is messing with me 😞😞              are you serious 😞😞              Let us go play baseball 🤔👤              This stupid grader is not working 😞😞              work is horrible 😞😞              Congratulation for having a baby 🤔😞              stop pissing me off 🤔😞              any suggestions for dinner 🤔👤              I love taking breaks 😞❤️              you brighten my day 😞😞              I boiled rice 🤔👤              she is a bully 😞😞              Why are you feeling bad 😞😞              I am upset 🤔😞              I am upset 🤔😞              give me the ball 🤔👤              My grandmother is the love of my life 😞❤️              enjoy your game 🤔👤              valentine day is near 😞😞              I miss you so much 😞❤️              throw the ball 🤔👤              My life is so boring 😞❤️              she said yes ❤️😞              will you be my valentine 😞😞              he can pitch really well 🤔👤              dance with me 🤔😞              I am hungry 😞👤              See you at the restaurant 🤔👤              I like to laugh 😞😞              I will run 😞👤              I like your jacket 😞❤️              i miss her 😞❤️              what is your favorite baseball game 🤔👤              Good job ❤️😞              I love you to the stars and back 🤔❤️              What you did was awesome 😞😞              ha ha ha lol 😞😞              I do not want to joke 😞😞              go away 🤔😞              yesterday we lost again 🤔😞              family is all I have 😞❤️              you are failing this exercise ❤️😞              Good joke ❤️😞              You deserve this nice prize ❤️😞              I did not have breakfast ❤️👤           </p>
-----------------	--

<p>(2) LSTM+SGD+50d</p>	<p>             I want to eat 🍴 🍴              he did not answer 😞 😞              he got a very nice raise 😊 😊              she got me a nice present ❤️ 😊              ha ha ha it was so funny 😊 😊              he is a good friend 😊 😊              I am upset 😞 😞              We had such a lovely dinner tonight 😊 😊              where is the food 🍴 🍴              Stop making this joke ha ha ha 😊 😊              where is the ball 🍴 🍴              work is hard 😞 😞              This girl is messing with me ❤️ 😞              are you serious 😞 😞              Let us go play baseball 🍴 🍴              This stupid grader is not working 😞 😞              work is horrible 😞 😞              Congratulation for having a baby 😊 😊              stop pissing me off 😞 😞              any suggestions for dinner 😞 🍴              I love taking breaks ❤️ ❤️              you brighten my day ❤️ 😊              I boiled rice 🍴 🍴              she is a bully 😞 😞              Why are you feeling bad 😞 😞              I am upset 😞 😞              give me the ball 🍴 🍴              My grandmother is the love of my life ❤️ ❤️              enjoy your game 🍴 🍴              valentine day is near 😊 😊              I miss you so much ❤️ ❤️              throw the ball 🍴 🍴              My life is so boring 😞 😞              she said yes 😊 😊              will you be my valentine 😊 😊              he can pitch really well 😊 🍴              dance with me 😊 😊              I am hungry 🍴 🍴              See you at the restaurant 😞 🍴              I like to laugh 😊 😊              I will run 🍴 🍴              I like your jacket ❤️ ❤️              i miss her 😊 ❤️              what is your favorite baseball game 🍴 🍴              Good job 😊 😊              I love you to the stars and back 😊 ❤️              What you did was awesome 😊 😊              ha ha ha lol 😊 😊              I do not want to joke 😞 😞              go away 😞 😞              yesterday we lost again 😞 😞              family is all I have ❤️ ❤️              you are failing this exercise 😞 😞              Good joke 😊 😊              You deserve this nice prize ❤️ 😊              I did not have breakfast 🍴 🍴           </p>
-------------------------	---

(3) LSTM+ADAM+50d	<p>             I want to eat 🍴 🍴              he did not answer 😞 😞              he got a very nice raise ❤️ 😊              she got me a nice present ❤️ 😊              ha ha ha it was so funny 😊 😊              he is a good friend 😊 😊              I am upset 😞 😞              We had such a lovely dinner tonight 🍴 😊              where is the food 🍴 🍴              Stop making this joke ha ha ha 😊 😊              where is the ball 🍴 🍴              work is hard 😞 😞              This girl is messing with me 😞 😞              are you serious 😞 😞              Let us go play baseball ❤️ 🍴              This stupid grader is not working 😞 😞              work is horrible 😞 😞              Congratulation for having a baby 🍴 😊              stop pissing me off 😞 😞              any suggestions for dinner 😊 🍴              I love taking breaks ❤️ ❤️              you brighten my day 😊 😊              I boiled rice 🍴 🍴              she is a bully 😞 😞              Why are you feeling bad 😞 😞              I am upset 😞 😞              give me the ball 🍴 🍴              My grandmother is the love of my life 😊 ❤️              enjoy your game 🍴 🍴              valentine day is near 😊 😊              I miss you so much ❤️ ❤️              throw the ball ❤️ 🍴              My life is so boring 😞 😞              she said yes 😊 😊              will you be my valentine 😊 😊              he can pitch really well 😊 🍴              dance with me 😊 😊              I am hungry 🍴 🍴              See you at the restaurant ❤️ 🍴              I like to laugh ❤️ 😊              I will run 🍴 🍴              I like your jacket ❤️ ❤️              i miss her 😊 ❤️              what is your favorite baseball game 🍴 🍴              Good job 😊 😊              I love you to the stars and back ❤️ ❤️              What you did was awesome 😊 😊              ha ha ha lol 😊 😊              I do not want to joke 😞 😞              go away ❤️ 😞              yesterday we lost again 😊 😊              family is all I have ❤️ ❤️              you are failing this exercise 😞 😞              Good joke ❤️ 😊              You deserve this nice prize 😊 😊              I did not have breakfast 🍴 🍴           </p>
-------------------	---

(4) LSTM+SGD+100d	<p>             I want to eat 🤔              he did not answer 😞              he got a very nice raise 😊              she got me a nice present ❤️              ha ha ha it was so funny 😊              he is a good friend ❤️              I am upset 😞              We had such a lovely dinner tonight 😊              where is the food 🤔              Stop making this joke ha ha ha 😊              where is the ball 🤔              work is hard 😞              This girl is messing with me 😞              are you serious 😞              Let us go play baseball 🤔              This stupid grader is not working 😞              work is horrible 😞              Congratulation for having a baby ❤️              stop pissing me off 🤔              any suggestions for dinner 🤔              I love taking breaks ❤️              you brighten my day 😊              I boiled rice 🤔              she is a bully ❤️              Why are you feeling bad 😞              I am upset 😞              give me the ball 🤔              My grandmother is the love of my life ❤️              enjoy your game 🤔              valentine day is near 🤔              I miss you so much ❤️              throw the ball 🤔              My life is so boring 🤔              she said yes 😊              will you be my valentine 😊              he can pitch really well 🤔              dance with me 😊              I am hungry ❤️              See you at the restaurant 🤔              I like to laugh 🤔              I will run 🤔              I like your jacket ❤️              i miss her ❤️              what is your favorite baseball game 🤔              Good job 😊              I love you to the stars and back 😊              What you did was awesome 😊              ha ha ha lol 😊              I do not want to joke 🤔              go away 🤔              yesterday we lost again 😞              family is all I have ❤️              you are failing this exercise 😞              Good joke 🤔              You deserve this nice prize 😊              I did not have breakfast ❤️           </p>
-------------------	---

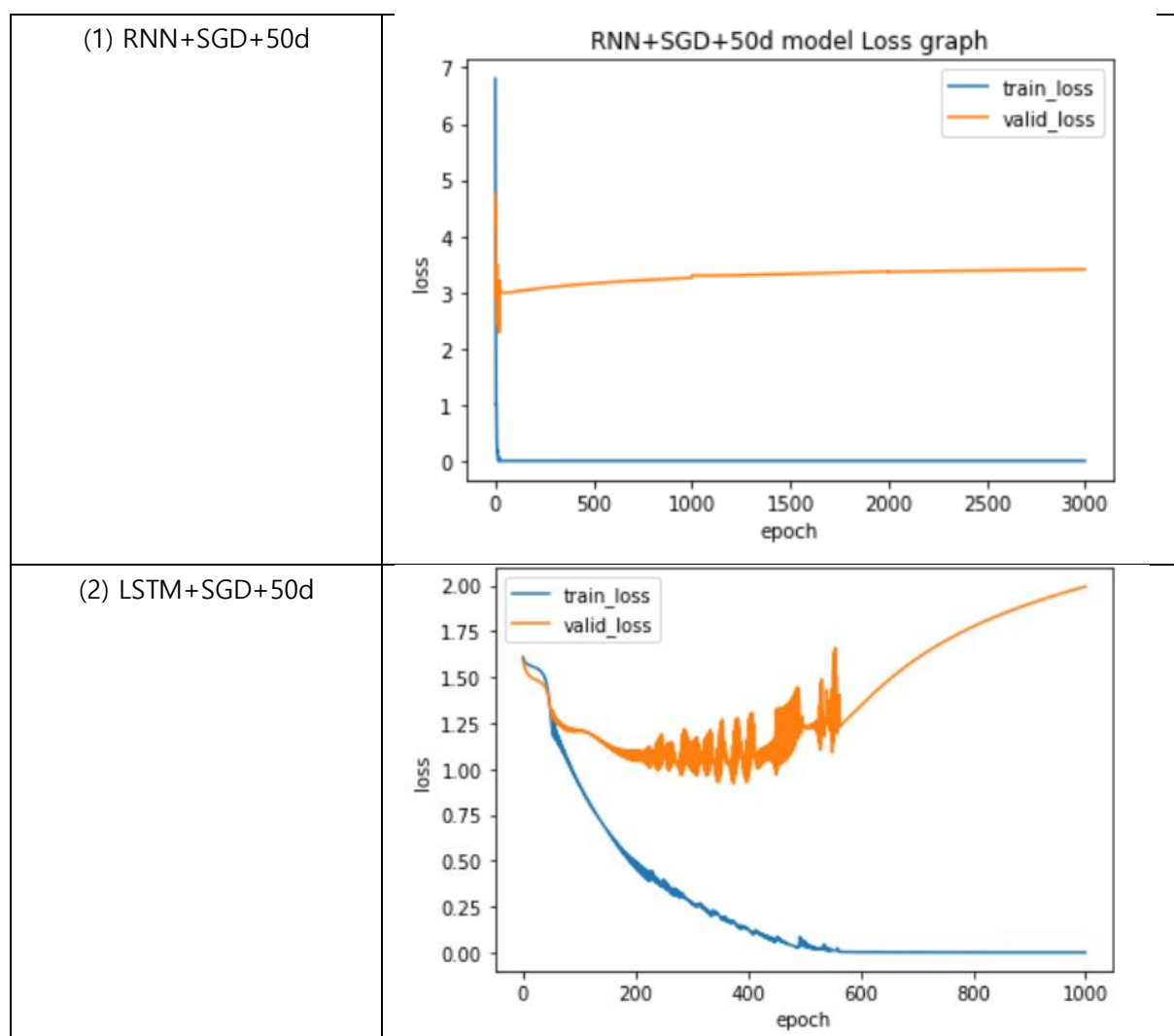
(5) LSTM+SGD+50d+dropout	<p>             I want to eat                he did not answer                he got a very nice raise                she got me a nice present                ha ha ha it was so funny                he is a good friend                I am upset                We had such a lovely dinner tonight                where is the food                Stop making this joke ha ha ha                where is the ball                work is hard                This girl is messing with me                are you serious                Let us go play baseball                This stupid grader is not working                work is horrible                Congratulation for having a baby                stop pissing me off                any suggestions for dinner                I love taking breaks                you brighten my day                I boiled rice                she is a bully                Why are you feeling bad                I am upset                give me the ball                My grandmother is the love of my life                enjoy your game                valentine day is near                I miss you so much                throw the ball                My life is so boring                she said yes                will you be my valentine                he can pitch really well                dance with me                I am hungry                See you at the restaurant                I like to laugh                I will run                I like your jacket                i miss her                what is your favorite baseball game                Good job                I love you to the stars and back                What you did was awesome                ha ha ha lol                I do not want to joke                go away                yesterday we lost again                family is all i have                you are failing this exercise                Good joke                You deserve this nice prize                I did not have breakfast   </p>
--------------------------	---

대부분의 모델의 출력 emoji가 슬픈표정인 것을 확인할 수 있다. 또한, RNN의 경우, test set에 대해 가장 좋은 성능이 33%밖에 되지 않기에, RNN의 경우는 하이퍼파라미터를 조정하더라도 총 test set(56개) 중 17~18개 정도만 정답을 맞췄다. 이와 다르게 LSTM은 RNN보다 높은 성능을 보였는데, 이는 RNN보다 LSTM이 똑 같은 sequence data를 입력받더라도 내부적으로 게이트를 뒤

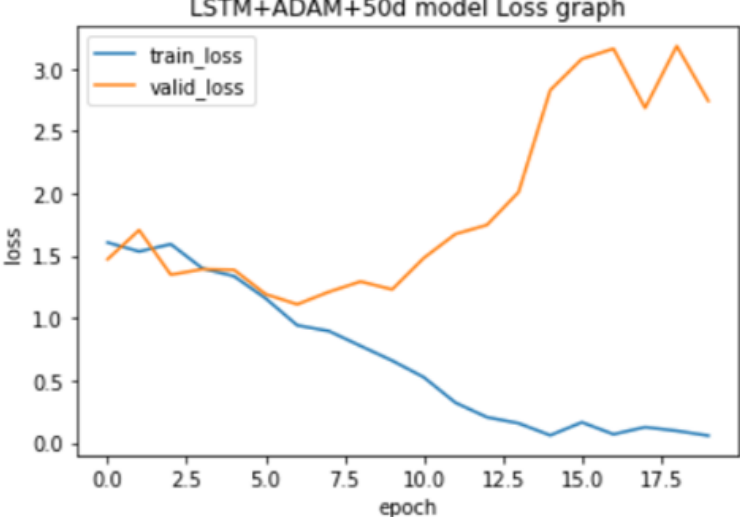
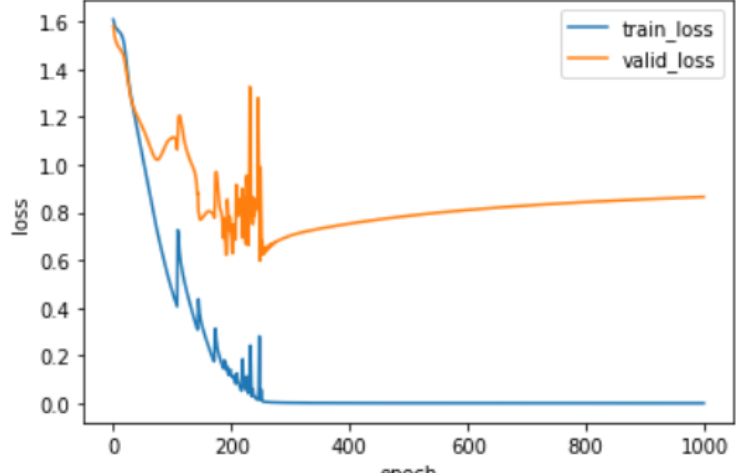
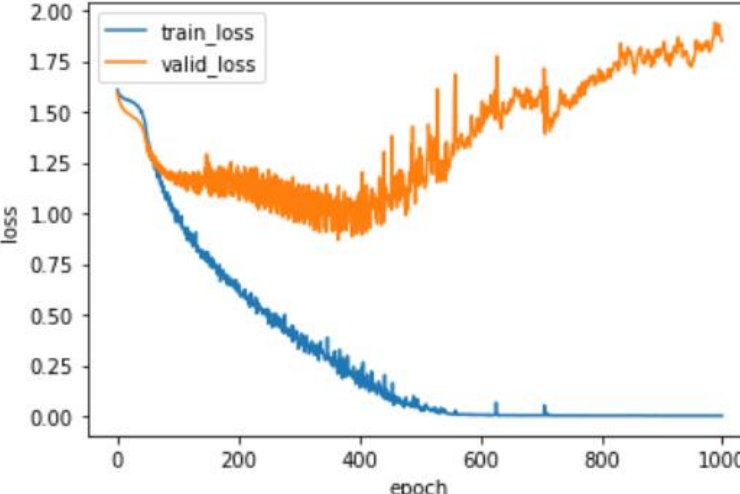
서 그 입력을 다르게 조작하므로써 이전 sequence data의 정보를 해당 입력에 맞춰 잘 조리하기 때문이다. 즉, RNN은 단순히 이전의 모든 데이터를 똑같이 처리하고 불러오지만, LSTM은 이전의 데이터를 모두 그대로 가져오지 않고 forget gate와 input gate, output gate 등을 이용해 자신의 현재 입력 sequence에 맞게 변화하여 이전데이터를 사용하기 때문이다.

### 3. Training Loss graph

1번 문항의 모델들의 loss graph이다. (1)번은 추가적인 epoch으로 인해 총 에폭수가 3000epoch이 된 것 외에 나머지 모델의 조건과 에폭 수는 1번 문항과 같다.





(3) LSTM+ADAM+50d	 <p>The graph shows training and validation loss over 18 epochs. Training loss (blue) decreases steadily from ~1.6 to ~0.1. Validation loss (orange) fluctuates initially, then rises sharply after epoch 10, peaking at ~3.2 around epoch 16 before dropping to ~2.8.</p>
(4) LSTM+SGD+100d	 <p>The graph shows training and validation loss over 1000 epochs. Training loss (blue) drops rapidly to near zero by epoch 200. Validation loss (orange) drops to ~0.6 at epoch 200, then gradually increases to ~0.85 by epoch 1000.</p>
(5) LSTM+SGD+50d+dropout	 <p>The graph shows training and validation loss over 1000 epochs. Training loss (blue) decreases steadily to near zero by epoch 600. Validation loss (orange) drops to ~1.1 at epoch 200, then increases steadily to ~1.9 by epoch 1000.</p>

(1) 모델의 경우, 1000epoch 이후 accuracy가 33% 밖에 되지 않아서, 추가적으로 2000 peoch을 더 돌린 총 3000epoch의 학습결과 그래프를 첨부하였다. 더 이상의 학습으로는 accuracy가 올라가지 않고 더 train set에 모델이 과적합될 뿐이었다.

(5) 모델의 경우, ppt에 나와있는 Dropout이 들어간 모델 개요와 똑같이 층을 구성하였다.  
(Embedding→LSTM→Dropout→LSTM→Dropout→Softmax)

Validation set에 대한 loss값이 epoch이 증가할수록 증가하는 경향을 보였기에 해당 모델은 train data에 과적합되었다고 볼 수 있다. 그러나 epoch수가 적은 경우(100~500 epoch), test set에 대한 accuracy가 20~30%정도만 나왔으며, epoch수가 1000이 되어야 test set에 대한 accuracy가 66%로 비교적 높게 나왔다. 따라서, 기본적인 LSTM 학습 시에는 sequential 한 data의 특성을 모델이 충분히 학습할 수 있도록 많은 train 과정이 필요함을 알 수 있다.

또한, Validation set의 경우 전체 train data set에서 20%로 따로 구성된 data set이지만, train data set이 132개로 validation set은 26개 밖에 되지 않는다. 따라서, 학습시키는 데이터 양도 train data는 106개로 더 적어지기 때문에 해당 데이터를 모델이 학습하게 되면 당연히 Overfitting이 일어날 수 밖에 없다.

전체적으로 validation set이 증가하는 경향을 보인다. 이는 train data가 적은 상태에서 에폭수를 높여갈수록 모델이 train data에 과적합되기 때문이다.

또한, SGD와 Adam을 비교했을 때, Adam의 경우 loss가 왔다갔다하는 구간이 덜하며, SGD는 Adam보다 loss값이 자주 왔다갔다하는 것을 볼 수 있다. 이는 Adam의 파라미터 갱신 과정은 그릇 바닥을 구르듯이 움직인다는 특성과 움직임의 좌우 흔들림이 적다는 특성과 관련하여 해석할 수 있다.

#### 4. Describe what Glove is & why we use Glove (~5 lines)

Glove는 말뭉치 전체의 통계 정보를 더 잘 반영하여 단어 벡터간 유사도 측정이 수월하게 반영하게 함으로써 기존의 Word2Vec의 '주변 단어 몇 개 내에서만 학습'하는 단점을 보완할 수 있는 임베딩 방법이다. 이번 과제에서는 이미 Glove 방식으로 학습된(pre-trained) 단어 임베딩 벡터(50d, 100d)들을 Embedding layer의 초기값으로 가져옴으로써 100여개의 '부족한 train data에서 단어들의 관계를 반영한 벡터로의 임베딩'을 시작했을 때보다 좋은 단어 임베딩 벡터를 얻을 수 있다.