

Emojify!

Welcome to the second assignment of Week 2. You are going to use word vector representations to build an Emojifier.

Have you ever wanted to make your text messages more expressive? Your emojifier app will help you do that. So rather than writing:

"Congratulations on the promotion! Let's get coffee and talk. Love you!"

The emojifier can automatically turn this into:

"Congratulations on the promotion! 👍 Let's get coffee and talk. ☕ Love you! ❤️"

- You will implement a model which inputs a sentence (such as "Let's go see the baseball game tonight!") and finds the most appropriate emoji to be used with this sentence (🌍).

Using word vectors to improve emoji lookups

- In many emoji interfaces, you need to remember that ❤️ is the "heart" symbol rather than the "love" symbol.
 - In other words, you'll have to remember to type "heart" to find the desired emoji, and typing "love" won't bring up that symbol.
- We can make a more flexible emoji interface by using word vectors!
- When using word vectors, you'll see that even if your training set explicitly relates only a few words to a particular emoji, your algorithm will be able to generalize and associate additional words in the test set to the same emoji.
 - This works even if those additional words don't even appear in the training set.
 - This allows you to build an accurate classifier mapping from sentences to emojis, even using a small training set.

What you'll build

1. In this exercise, you'll start with a baseline model (Emojifier-V1) using word embeddings.
2. Then you will build a more sophisticated model (Emojifier-V2) that further incorporates an LSTM.

Updates

If you were working on the notebook before this update...

- The current notebook is version "2a".
- You can find your original work saved in the notebook with the previous version name ("v2")
- To view the file directory, go to the menu "File->Open", and this will open a new tab that shows the file directory.

List of updates

- sentence_to_avg
 - Updated instructions.
 - Use separate variables to store the total and the average (instead of just avg).
 - Additional hint about how to initialize the shape of avg vector.
- sentences_to_indices
 - Updated preceding text and instructions, added additional hints.
- pretrained_embedding_layer
 - Additional instructions to explain how to implement each step.
- Emoify_V2
 - Modifies instructions to specify which parameters are needed for each Keras layer.
 - Remind users of Keras syntax.
 - Explanation of how to use the layer object that is returned by pretrained_embedding_layer.
 - Provides sample Keras code.

Let's get started! Run the following cell to load the package you are going to use.

```
In [1]: import numpy as np
        from emo_utils import *
        import emoji
        import matplotlib.pyplot as plt

        %matplotlib inline
```

1 - Baseline model: Emojifier-V1






1.1 - Dataset EMOJISSET

Let's start by building a simple baseline classifier.

You have a tiny dataset (X, Y) where:

- X contains 127 sentences (strings).
- Y contains an integer label between 0 and 4 corresponding to an emoji for each sentence.

X (sentences)	Y (labels)
I love you	0
Congrats on the new job	2
I think I will end up alone	3
I want to have sushi for dinner!	4
It was funny lol	2
she did not answer my text	3
Happy new year	2
my algorithm performs poorly	3
he can pitch really well	1
you are failing this exercise	3
you did well on your exam.	2
What you did was awesome	2
I am frustrated	3

code	emoji	label
:heart:		0
:baseball:		1
:smile:		2
:disappointed:		3
:fork_and_knife:		4

****Figure 1**:** EMOJISSET - a classification problem with 5 classes. A few examples of sentences are given here.

Let's load the dataset using the code below. We split the dataset between training (127 examples) and testing (56 examples).

```
In [2]: X_train, Y_train = read_csv('data/train_emoji.csv')
        X_test, Y_test = read_csv('data/tesss.csv')
```

```
In [3]: maxLen = len(max(X_train, key=len).split())
```

Run the following cell to print sentences from X_train and corresponding labels from Y_train.

- Change idx to see different examples.
- Note that due to the font used by iPython notebook, the heart emoji may be colored black rather than red.

```
In [4]: for idx in range(10):
        print(X_train[idx], label_to_emoji(Y_train[idx]))
```

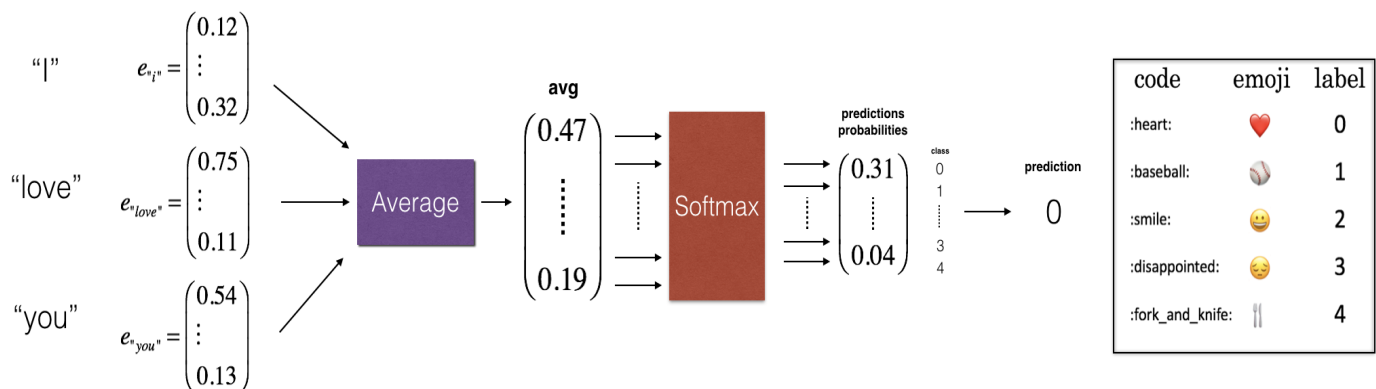
```
never talk to me again 😞
I am proud of your achievements 😊
It is the worst day in my life 😞
Miss you so much ❤️
food is life 🍴
I love you mum ❤️
Stop saying bullshit 😞
congratulations on your acceptance 😊
The assignment is too long 😞
I want to play 🏈
```

1.2 - Overview of the Emojiifier-V1

In this part, you are going to implement a baseline model called "Emojiifier-v1".

Emojiifier - V1

$$\sigma(We+b) \rightarrow 0.74 \xrightarrow{>0.5} 1$$



****Figure 2****: Baseline model (Emojiifier-V1).

Inputs and outputs

- The input of the model is a string corresponding to a sentence (e.g. "I love you").
- The output will be a probability vector of shape (1,5), (there are 5 emojis to choose from).
- The (1,5) probability vector is passed to an argmax layer, which extracts the index of the emoji with the highest probability.

One-hot encoding

- To get our labels into a format suitable for training a softmax classifier, let's convert Y from its current shape $(m, 1)$ into a "one-hot representation" $(m, 5)$,
 - Each row is a one-hot vector giving the label of one example.
 - Here, Y_{oh} stands for "Y-one-hot" in the variable names $Y_{\text{oh_train}}$ and $Y_{\text{oh_test}}$:

```
In [5]: Y_oh_train = convert_to_one_hot(Y_train, C = 5)
        Y_oh_test = convert_to_one_hot(Y_test, C = 5)
```

Let's see what `convert_to_one_hot()` did. Feel free to change index to print out different values.

```
In [6]: idx = 50
        print(f"Sentence '{X_train[50]}' has label index {Y_train[idx]}, which is emoji {label_to_emoji(Y_train[idx])}", )
        print(f"Label index {Y_train[idx]} in one-hot encoding format is {Y_oh_train[idx]}")
```

```
Sentence 'I missed you' has label index 0, which is emoji 🍷
Label index 0 in one-hot encoding format is [ 1.  0.  0.  0.  0.]
```

All the data is now ready to be fed into the Emojify-V1 model. Let's implement the model!

1.3 - Implementing Emojifier-V1

As shown in Figure 2 (above), the first step is to:

- Convert each word in the input sentence into their word vector representations.
- Then take an average of the word vectors.
- Similar to the previous exercise, we will use pre-trained 50-dimensional GloVe embeddings.

Run the following cell to load the `word_to_vec_map`, which contains all the vector representations.

```
In [7]: word_to_index, index_to_word, word_to_vec_map = read_glove_vecs('../readonl
        y/glove.6B.50d.txt')
```

You've loaded:

- `word_to_index`: dictionary mapping from words to their indices in the vocabulary
 - (400,001 words, with the valid indices ranging from 0 to 400,000)
- `index_to_word`: dictionary mapping from indices to their corresponding words in the vocabulary
- `word_to_vec_map`: dictionary mapping words to their GloVe vector representation.

Run the following cell to check if it works.

```
In [8]: word = "cucumber"
        idx = 289846
        print("the index of", word, "in the vocabulary is", word_to_index[word])
        print("the", str(idx) + "th word in the vocabulary is", index_to_word[idx])
```

```
the index of cucumber in the vocabulary is 113317
the 289846th word in the vocabulary is potatoes
```

Exercise: Implement `sentence_to_avg()`. You will need to carry out two steps:

1. Convert every sentence to lower-case, then split the sentence into a list of words.
 - `X.lower()` and `X.split()` might be useful.
2. For each word in the sentence, access its GloVe representation.
 - Then take the average of all of these word vectors.
 - You might use `numpy.zeros()`.

Additional Hints

- When creating the avg array of zeros, you'll want it to be a vector of the same shape as the other word vectors in the `word_to_vec_map`.
 - You can choose a word that exists in the `word_to_vec_map` and access its `.shape` field.
 - Be careful not to hard code the word that you access. In other words, don't assume that if you see the word 'the' in the `word_to_vec_map` within this notebook, that this word will be in the `word_to_vec_map` when the function is being called by the automatic grader.
 - Hint: you can use any one of the word vectors that you retrieved from the input sentence to find the shape of a word vector.

```
In [11]: # GRADED FUNCTION: sentence_to_avg

def sentence_to_avg(sentence, word_to_vec_map):
    """
    Converts a sentence (string) into a list of words (strings). Extracts the
    GloVe representation of each word
    and averages its value into a single vector encoding the meaning of the se
    ntence.

    Arguments:
    sentence -- string, one training example from X
    word_to_vec_map -- dictionary mapping every word in a vocabulary into its
    50-dimensional vector representation

    Returns:
    avg -- average vector encoding information about the sentence, numpy-array
    of shape (50,)
    """

    ### START CODE HERE ###
    # Step 1: Split sentence into list of lower case words (~ 1 line)
    words = sentence.lower().split()

    # Initialize the average word vector, should have the same shape as your w
    ord vectors.
    avg = np.zeros(word_to_vec_map[words[0]].shape)

    # Step 2: average the word vectors. You can loop over the words in the lis
    t "words".
    total = 0
    for w in words:
        total += word_to_vec_map[w]
    avg = total / len(words)

    ### END CODE HERE ###

    return avg
```

```
In [12]: avg = sentence_to_avg("Morrocan couscous is my favorite dish", word_to_vec_map)
print("avg = \n", avg)
```

```
avg =
[ -0.008005   0.56370833 -0.50427333  0.258865    0.55131103  0.03104983
 -0.21013718  0.16893933 -0.09590267  0.141784   -0.15708967  0.18525867
  0.6495785   0.38371117  0.21102167  0.11301667  0.02613967  0.26037767
  0.05820667 -0.01578167 -0.12078833 -0.02471267  0.4128455   0.5152061
  0.38756167 -0.898661   -0.535145    0.33501167  0.68806933 -0.2156265
  1.797155    0.10476933 -0.36775333  0.750785    0.10282583  0.348925
 -0.27262833  0.66768    -0.10706167 -0.283635    0.59580117  0.28747333
 -0.3366635   0.23393817  0.34349183  0.178405    0.1166155   -0.076433
  0.1445417   0.09808667]
```

Expected Output:

```

avg =
[ -0.008005    0.56370833 -0.50427333  0.258865    0.55131103  0.03104983
 -0.21013718  0.16893933 -0.09590267  0.141784   -0.15708967  0.18525867
  0.6495785   0.38371117  0.21102167  0.11301667  0.02613967  0.26037767
  0.05820667 -0.01578167 -0.12078833 -0.02471267  0.4128455   0.5152061
  0.38756167 -0.898661   -0.535145   0.33501167  0.68806933 -0.2156265
  1.797155    0.10476933 -0.36775333  0.750785    0.10282583  0.348925
 -0.27262833  0.66768    -0.10706167 -0.283635    0.59580117  0.28747333
 -0.3366635   0.23393817  0.34349183  0.178405    0.1166155   -0.076433
  0.1445417   0.09808667]

```

Model

You now have all the pieces to finish implementing the `model()` function. After using `sentence_to_avg()` you need to:

- Pass the average through forward propagation
- Compute the cost
- Backpropagate to update the softmax parameters

Exercise: Implement the `model()` function described in Figure (2).

- The equations you need to implement in the forward pass and to compute the cross-entropy cost are below:
- The variable Y_{oh} ("Y one hot") is the one-hot encoding of the output labels.

$$\begin{aligned}
 z^{(i)} &= W \cdot avg^{(i)} + b \\
 a^{(i)} &= softmax(z^{(i)}) \\
 \mathcal{L}^{(i)} &= - \sum_{k=0}^{n_y-1} Y_{oh,k}^{(i)} * \log(a_k^{(i)})
 \end{aligned}$$

Note It is possible to come up with a more efficient vectorized implementation. For now, let's use nested for loops to better understand the algorithm, and for easier debugging.

We provided the function `softmax()`, which was imported earlier.

In [28]: # GRADED FUNCTION: model

```

def model(X, Y, word_to_vec_map, learning_rate = 0.01, num_iterations = 400):
    """
    Model to train word vector representations in numpy.

    Arguments:
    X -- input data, numpy array of sentences as strings, of shape (m, 1)
    Y -- labels, numpy array of integers between 0 and 7, numpy-array of shape
    (m, 1)
    word_to_vec_map -- dictionary mapping every word in a vocabulary into its
    50-dimensional vector representation
    learning_rate -- learning_rate for the stochastic gradient descent algorit
    hm
    num_iterations -- number of iterations

    Returns:
    pred -- vector of predictions, numpy-array of shape (m, 1)
    W -- weight matrix of the softmax layer, of shape (n_y, n_h)
    b -- bias of the softmax layer, of shape (n_y,)
    """

    np.random.seed(1)

    # Define number of training examples
    m = Y.shape[0]                # number of training examples
    n_y = 5                       # number of classes
    n_h = 50                      # dimensions of the GloVe vectors

    # Initialize parameters using Xavier initialization
    W = np.random.randn(n_y, n_h) / np.sqrt(n_h)
    b = np.zeros((n_y,))

    # Convert Y to Y_onehot with n_y classes
    Y_oh = convert_to_one_hot(Y, C = n_y)

    # Optimization Loop
    for t in range(num_iterations): # Loop over the number of iterations
        for i in range(m):         # Loop over the training examples
            ### START CODE HERE ### (= 4 lines of code)
            # Average the word vectors of the words from the i'th training exa
            mple

            avg = sentence_to_avg(X[i], word_to_vec_map)

            # Forward propagate the avg through the softmax layer
            z = np.dot(W, avg) + b
            a = softmax(z)

            # Compute cost using the i'th training label's one hot representat
            ion and "A" (the output of the softmax)
            cost = - np.sum(Y_oh[i] * np.log(a))
            ### END CODE HERE ###

            # Compute gradients
            dz = a - Y_oh[i]
            dW = np.dot(dz.reshape(n_y,1), avg.reshape(1, n_h))

```

```

        db = dz

        # Update parameters with Stochastic Gradient Descent
        W = W - learning_rate * dW
        b = b - learning_rate * db

        if t % 100 == 0:
            print("Epoch: " + str(t) + " --- cost = " + str(cost))
            pred = predict(X, Y, W, b, word_to_vec_map) #predict is defined in
emo_utils.py

    return pred, W, b

```

```

In [29]: print(X_train.shape)
print(Y_train.shape)
print(np.eye(5)[Y_train.reshape(-1)].shape)
print(X_train[0])
print(type(X_train))
Y = np.asarray([5,0,0,5, 4, 4, 4, 6, 6, 4, 1, 1, 5, 6, 6, 3, 6, 3, 4, 4])
print(Y.shape)

X = np.asarray(['I am going to the bar tonight', 'I love you', 'miss you my de
ar',
'Lets go party and drinks', 'Congrats on the new job', 'Congratulations',
'I am so happy for you', 'Why are you feeling bad', 'What is wrong with you',
'You totally deserve this prize', 'Let us go play football',
'Are you down for football this afternoon', 'Work hard play harder',
'It is suprising how people can be dumb sometimes',
'I am very disappointed', 'It is the best day in my life',
'I think I will end up alone', 'My life is so boring', 'Good job',
'Great so awesome'])

print(X.shape)
print(np.eye(5)[Y_train.reshape(-1)].shape)
print(type(X_train))

(132,)
(132,)
(132, 5)
never talk to me again
<class 'numpy.ndarray'>
(20,)
(20,)
(132, 5)
<class 'numpy.ndarray'>

```

Run the next cell to train your model and learn the softmax parameters (W,b).

```
In [30]: pred, W, b = model(X_train, Y_train, word_to_vec_map)
         print(pred)
```

```
Epoch: 0 --- cost = 1.95204988128
Accuracy: 0.348484848485
Epoch: 100 --- cost = 0.0797181872601
Accuracy: 0.931818181818
Epoch: 200 --- cost = 0.0445636924368
Accuracy: 0.954545454545
Epoch: 300 --- cost = 0.0343226737879
Accuracy: 0.969696969697
[[ 3.]
 [ 2.]
 [ 3.]
 [ 0.]
 [ 4.]
 [ 0.]
 [ 3.]
 [ 2.]
 [ 3.]
 [ 1.]
 [ 3.]
 [ 3.]
 [ 1.]
 [ 3.]
 [ 2.]
 [ 3.]
 [ 2.]
 [ 3.]
 [ 1.]
 [ 2.]
 [ 3.]
 [ 0.]
 [ 2.]
 [ 2.]
 [ 2.]
 [ 1.]
 [ 4.]
 [ 3.]
 [ 3.]
 [ 4.]
 [ 0.]
 [ 3.]
 [ 4.]
 [ 2.]
 [ 0.]
 [ 3.]
 [ 2.]
 [ 2.]
 [ 3.]
 [ 4.]
 [ 2.]
 [ 2.]
 [ 0.]
 [ 2.]
 [ 3.]
 [ 0.]
 [ 3.]
 [ 2.]
 [ 4.]
```

[3.]
[0.]
[3.]
[3.]
[3.]
[4.]
[2.]
[1.]
[1.]
[1.]
[2.]
[3.]
[1.]
[0.]
[0.]
[0.]
[3.]
[4.]
[4.]
[2.]
[2.]
[1.]
[2.]
[0.]
[3.]
[2.]
[2.]
[0.]
[3.]
[3.]
[1.]
[2.]
[1.]
[2.]
[2.]
[4.]
[3.]
[3.]
[2.]
[4.]
[0.]
[0.]
[3.]
[3.]
[3.]
[3.]
[2.]
[0.]
[1.]
[2.]
[3.]
[0.]
[2.]
[2.]
[2.]
[3.]
[2.]

```
[ 2.]
[ 2.]
[ 4.]
[ 1.]
[ 1.]
[ 3.]
[ 3.]
[ 4.]
[ 1.]
[ 2.]
[ 1.]
[ 1.]
[ 3.]
[ 1.]
[ 0.]
[ 4.]
[ 0.]
[ 3.]
[ 3.]
[ 4.]
[ 4.]
[ 1.]
[ 4.]
[ 3.]
[ 0.]
[ 2.]]
```

Expected Output (on a subset of iterations):

Epoch: 0	cost = 1.95204988128	Accuracy: 0.348484848485
Epoch: 100	cost = 0.0797181872601	Accuracy: 0.931818181818
Epoch: 200	cost = 0.0445636924368	Accuracy: 0.954545454545
Epoch: 300	cost = 0.0343226737879	Accuracy: 0.969696969697

Great! Your model has pretty high accuracy on the training set. Lets now see how it does on the test set.

1.4 - Examining test set performance

- Note that the predict function used here is defined in emo_util.spy.

```
In [31]: print("Training set:")
pred_train = predict(X_train, Y_train, W, b, word_to_vec_map)
print('Test set:')
pred_test = predict(X_test, Y_test, W, b, word_to_vec_map)
```

```
Training set:
Accuracy: 0.977272727273
Test set:
Accuracy: 0.857142857143
```

Expected Output:


Train set accuracy	97.7
Test set accuracy	85.7

- Random guessing would have had 20% accuracy given that there are 5 classes. ($1/5 = 20\%$).
- This is pretty good performance after training on only 127 examples.

The model matches emojis to relevant words

In the training set, the algorithm saw the sentence

"I love you"







with the label .

- You can check that the word "adore" does not appear in the training set.
- Nonetheless, lets see what happens if you write "I adore you."

```
In [37]: X_my_sentences = np.array(["i want you", "i love you", "funny lol", "lets play
with a ball", "food is ready", "not feeling happy"])
Y_my_labels = np.array([[0], [0], [2], [1], [4],[3]])

pred = predict(X_my_sentences, Y_my_labels , W, b, word_to_vec_map)
print_predictions(X_my_sentences, pred)
```

Accuracy: 0.833333333333

i want you 
i love you 
funny lol 
lets play with a ball 
food is ready 
not feeling happy 

Amazing!

- Because *adore* has a similar embedding as *love*, the algorithm has generalized correctly even to a word it has never seen before.
- Words such as *heart*, *dear*, *beloved* or *adore* have embedding vectors similar to *love*.
 - Feel free to modify the inputs above and try out a variety of input sentences.
 - How well does it work?

Word ordering isn't considered in this model

- Note that the model doesn't get the following sentence correct:

"not feeling happy"

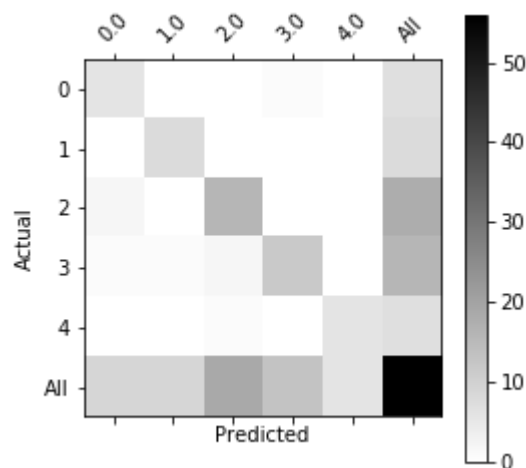
- This algorithm ignores word ordering, so is not good at understanding phrases like "not happy."

Confusion matrix

- Printing the confusion matrix can also help understand which classes are more difficult for your model.
- A confusion matrix shows how often an example whose label is one class ("actual" class) is mislabeled by the algorithm with a different class ("predicted" class).


```
In [33]: print(Y_test.shape)
print('      '+ label_to_emoji(0)+ '      ' + label_to_emoji(1) + '      ' +
label_to_emoji(2)+ '      ' + label_to_emoji(3)+ '      ' + label_to_emoji(4))
print(pd.crosstab(Y_test, pred_test.reshape(56,), rownames=['Actual'], colname
s=['Predicted'], margins=True))
plot_confusion_matrix(Y_test, pred_test)
```

```
(56,)
      ♡   🤔   😊   😞   🙄
Predicted 0.0 1.0 2.0 3.0 4.0 All
Actual
0         6   0   0   1   0   7
1         0   8   0   0   0   8
2         2   0  16   0   0  18
3         1   1   2  12   0  16
4         0   0   1   0   6   7
All        9   9  19  13   6  56
```



What you should remember from this section

- Even with a 127 training examples, you can get a reasonably good model for Emojifying.
 - This is due to the generalization power word vectors gives you.
- Emojify-V1 will perform poorly on sentences such as *"This movie is not good and not enjoyable"*
 - It doesn't understand combinations of words.
 - It just averages all the words' embedding vectors together, without considering the ordering of words.

You will build a better algorithm in the next section!

2 - Emojifier-V2: Using LSTMs in Keras:

Let's build an LSTM model that takes word **sequences** as input!

- This model will be able to account for the word ordering.
- Emojifier-V2 will continue to use pre-trained word embeddings to represent words.
- We will feed word embeddings into an LSTM.
- The LSTM will learn to predict the most appropriate emoji.

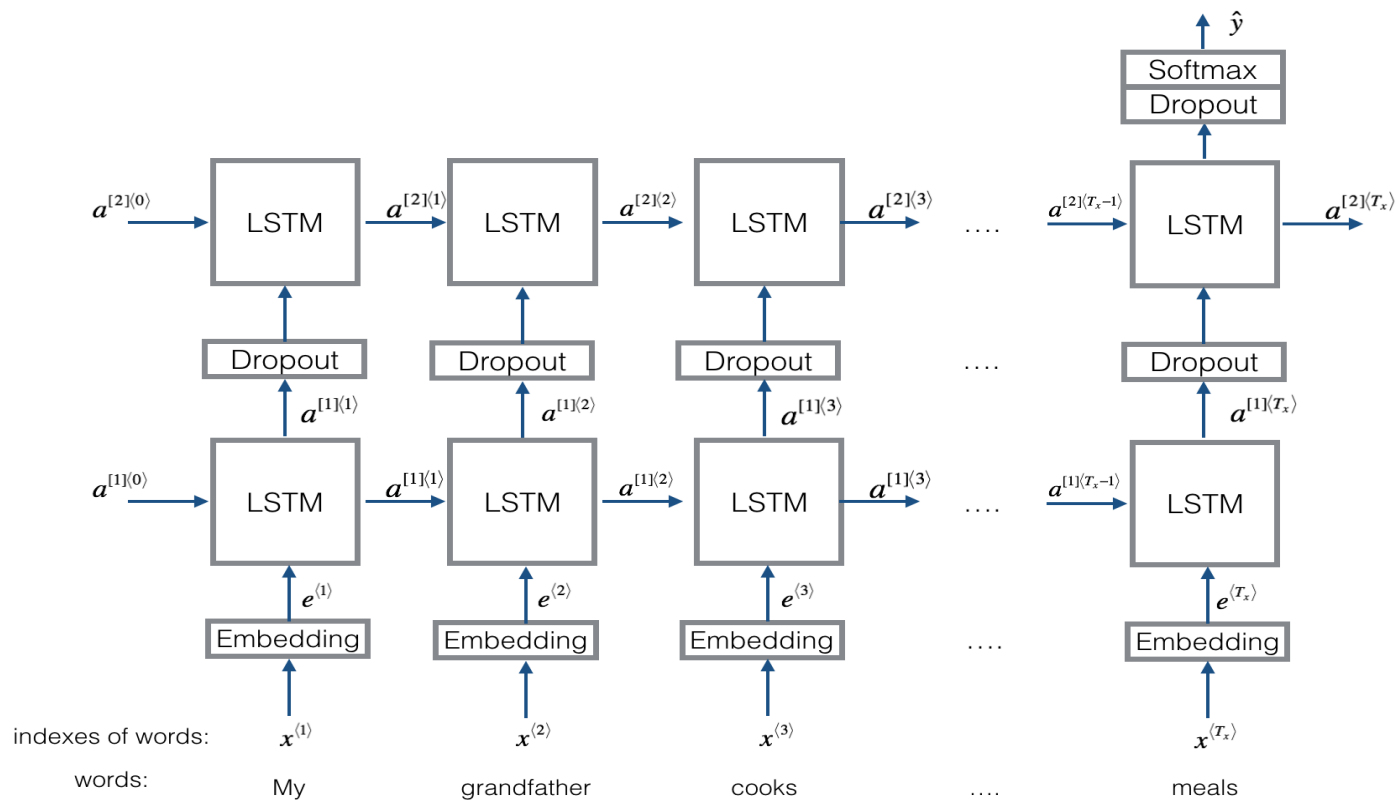
Run the following cell to load the Keras packages.

```
In [38]: import numpy as np
np.random.seed(0)
from keras.models import Model
from keras.layers import Dense, Input, Dropout, LSTM, Activation
from keras.layers.embeddings import Embedding
from keras.preprocessing import sequence
from keras.initializers import glorot_uniform
np.random.seed(1)
```

Using TensorFlow backend.

2.1 - Overview of the model

Here is the Emojifier-v2 you will implement:



****Figure 3**:** Emojifier-V2. A 2-layer LSTM sequence classifier.

2.2 Keras and mini-batching

- In this exercise, we want to train Keras using mini-batches.
- However, most deep learning frameworks require that all sequences in the same mini-batch have the **same length**.
 - This is what allows vectorization to work: If you had a 3-word sentence and a 4-word sentence, then the computations needed for them are different (one takes 3 steps of an LSTM, one takes 4 steps) so it's just not possible to do them both at the same time.

Padding handles sequences of varying length

- The common solution to handling sequences of **different length** is to use padding. Specifically:
 - Set a maximum sequence length
 - Pad all sequences to have the same length.

Example of padding

- Given a maximum sequence length of 20, we could pad every sentence with "0"s so that each input sentence is of length 20.
- Thus, the sentence "I love you" would be represented as $(e_I, e_{love}, e_{you}, \vec{0}, \vec{0}, \dots, \vec{0})$.
- In this example, any sentences longer than 20 words would have to be truncated.
- One way to choose the maximum sequence length is to just pick the length of the longest sentence in the training set.

2.3 - The Embedding layer

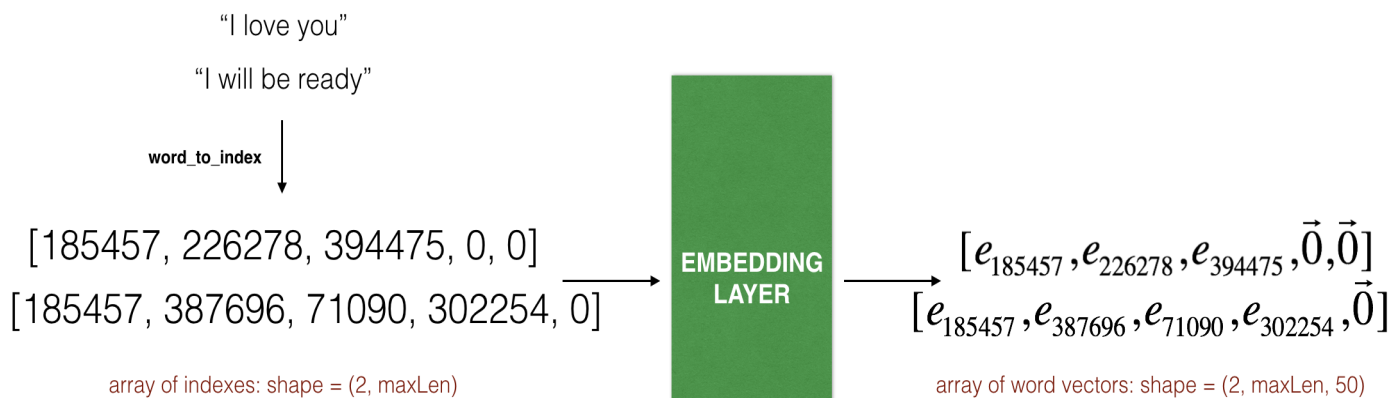
- In Keras, the embedding matrix is represented as a "layer".
- The embedding matrix maps word indices to embedding vectors.
 - The word indices are positive integers.
 - The embedding vectors are dense vectors of fixed size.
 - When we say a vector is "dense", in this context, it means that most of the values are non-zero. As a counter-example, a one-hot encoded vector is not "dense."
- The embedding matrix can be derived in two ways:
 - Training a model to derive the embeddings from scratch.
 - Using a pretrained embedding

Using and updating pre-trained embeddings

- In this part, you will learn how to create an `Embedding()` (<https://keras.io/layers/embeddings/>) layer in Keras
- You will initialize the Embedding layer with the GloVe 50-dimensional vectors.
- In the code below, we'll show you how Keras allows you to either train or leave fixed this layer.
- Because our training set is quite small, we will leave the GloVe embeddings fixed instead of updating them.

Inputs and outputs to the embedding layer

- The Embedding() layer's input is an integer matrix of size **(batch size, max input length)**.
 - This input corresponds to sentences converted into lists of indices (integers).
 - The largest integer (the highest word index) in the input should be no larger than the vocabulary size.
- The embedding layer outputs an array of shape (batch size, max input length, dimension of word vectors).
- The figure shows the propagation of two example sentences through the embedding layer.
 - Both examples have been zero-padded to a length of max_len=5.
 - The word embeddings are 50 units in length.
 - The final dimension of the representation is (2, max_len, 50).



****Figure 4****: Embedding layer

Prepare the input sentences

Exercise:

- Implement sentences_to_indices, which processes an array of sentences (X) and returns inputs to the embedding layer:
 - Convert each training sentences into a list of indices (the indices correspond to each word in the sentence)
 - Zero-pad all these lists so that their length is the length of the longest sentence.

Additional Hints

- Note that you may have considered using the enumerate() function in the for loop, but for the purposes of passing the autograder, please follow the starter code by initializing and incrementing j explicitly.

```
In [ ]: for idx, val in enumerate(["I", "like", "learning"]):
        print(idx, val)
```

```

In [41]: # GRADED FUNCTION: sentences_to_indices

def sentences_to_indices(X, word_to_index, max_len):
    """
    Converts an array of sentences (strings) into an array of indices corresponding to words in the sentences.
    The output shape should be such that it can be given to `Embedding()` (described in Figure 4).

    Arguments:
    X -- array of sentences (strings), of shape (m, 1)
    word_to_index -- a dictionary containing the each word mapped to its index
    max_len -- maximum number of words in a sentence. You can assume every sentence in X is no longer than this.

    Returns:
    X_indices -- array of indices corresponding to words in the sentences from X, of shape (m, max_len)
    """

    m = X.shape[0]                                # number of training examples

    ### START CODE HERE ###
    # Initialize X_indices as a numpy matrix of zeros and the correct shape (= 1 line)
    X_indices = np.zeros((m, max_len))

    for i in range(m):                            # Loop over training examples

        # Convert the ith training sentence in lower case and split it into words. You should get a list of words.
        sentence_words = X[i].lower().split()

        # Initialize j to 0
        j = 0

        # Loop over the words of sentence_words
        for w in sentence_words:
            # Set the (i,j)th entry of X_indices to the index of the correct word.
            X_indices[i, j] = word_to_index[w]
            # Increment j to j + 1
            j = j + 1

    ### END CODE HERE ###

    return X_indices

```

Run the following cell to check what `sentences_to_indices()` does, and check your results.

```
In [42]: X1 = np.array(["funny lol", "lets play baseball", "food is ready for you"])
X1_indices = sentences_to_indices(X1,word_to_index, max_len = 5)
print("X1 =", X1)
print("X1_indices =\n", X1_indices)
```

```
X1 = ['funny lol' 'lets play baseball' 'food is ready for you']
X1_indices =
[[ 155345.  225122.         0.         0.         0.]
 [ 220930.  286375.   69714.         0.         0.]
 [ 151204.  192973.  302254.  151349.  394475.]]
```

Expected Output:

```
X1 = ['funny lol' 'lets play baseball' 'food is ready for you']
X1_indices =
[[ 155345.  225122.         0.         0.         0.]
 [ 220930.  286375.   69714.         0.         0.]
 [ 151204.  192973.  302254.  151349.  394475.]]
```

Build embedding layer

- Let's build the `Embedding()` layer in Keras, using pre-trained word vectors.
- The embedding layer takes as input a list of word indices.
 - `sentences_to_indices()` creates these word indices.
- The embedding layer will return the word embeddings for a sentence.

Exercise: Implement `pretrained_embedding_layer()` with these steps:

1. Initialize the embedding matrix as a numpy array of zeros.
 - The embedding matrix has a row for each unique word in the vocabulary.
 - There is one additional row to handle "unknown" words.
 - So `vocab_len` is the number of unique words plus one.
 - Each row will store the vector representation of one word.
 - For example, one row may be 50 positions long if using GloVe word vectors.
 - In the code below, `emb_dim` represents the length of a word embedding.
2. Fill in each row of the embedding matrix with the vector representation of a word
 - Each word in `word_to_index` is a string.
 - `word_to_vec_map` is a dictionary where the keys are strings and the values are the word vectors.
3. Define the Keras embedding layer.
 - Use `Embedding()` (<https://keras.io/layers/embeddings/>).
 - The input dimension is equal to the vocabulary length (number of unique words plus one).
 - The output dimension is equal to the number of positions in a word embedding.
 - Make this layer's embeddings fixed.
 - If you were to set `trainable = True`, then it will allow the optimization algorithm to modify the values of the word embeddings.
 - In this case, we don't want the model to modify the word embeddings.
4. Set the embedding weights to be equal to the embedding matrix.
 - Note that this is part of the code is already completed for you and does not need to be modified.


```

In [45]: # GRADED FUNCTION: pretrained_embedding_layer

def pretrained_embedding_layer(word_to_vec_map, word_to_index):
    """
    Creates a Keras Embedding() layer and loads in pre-trained GloVe 50-dimensional vectors.

    Arguments:
    word_to_vec_map -- dictionary mapping words to their GloVe vector representation.
    word_to_index -- dictionary mapping from words to their indices in the vocabulary (400,001 words)

    Returns:
    embedding_layer -- pretrained layer Keras instance
    """

    vocab_len = len(word_to_index) + 1 # adding 1 to fit Keras embedding (requirement)
    emb_dim = word_to_vec_map["cucumber"].shape[0] # define dimensionality of your GloVe word vectors (= 50)

    ### START CODE HERE ###
    # Step 1
    # Initialize the embedding matrix as a numpy array of zeros.
    # See instructions above to choose the correct shape.
    emb_matrix = np.zeros((vocab_len, emb_dim))

    # Step 2
    # Set each row "idx" of the embedding matrix to be
    # the word vector representation of the idx'th word of the vocabulary
    for word, idx in word_to_index.items():
        emb_matrix[idx, :] = word_to_vec_map[word]

    # Step 3
    # Define Keras embedding layer with the correct input and output sizes
    # Make it non-trainable.
    embedding_layer = Embedding(vocab_len, emb_dim, trainable=False)
    ### END CODE HERE ###

    # Step 4 (already done for you; please do not modify)
    # Build the embedding layer, it is required before setting the weights of the embedding layer.
    embedding_layer.build((None,)) # Do not modify the "None". This line of code is complete as-is.

    # Set the weights of the embedding layer to the embedding matrix. Your layer is now pretrained.
    embedding_layer.set_weights([emb_matrix])

    return embedding_layer

```

```
In [46]: embedding_layer = pretrained_embedding_layer(word_to_vec_map, word_to_index)
         print("weights[0][1][3] =", embedding_layer.get_weights()[0][1][3])

         weights[0][1][3] = -0.3403
```

Expected Output:

```
weights[0][1][3] = -0.3403
```

2.3 Building the Emojiifier-V2

Lets now build the Emojiifier-V2 model.

- You feed the embedding layer's output to an LSTM network.

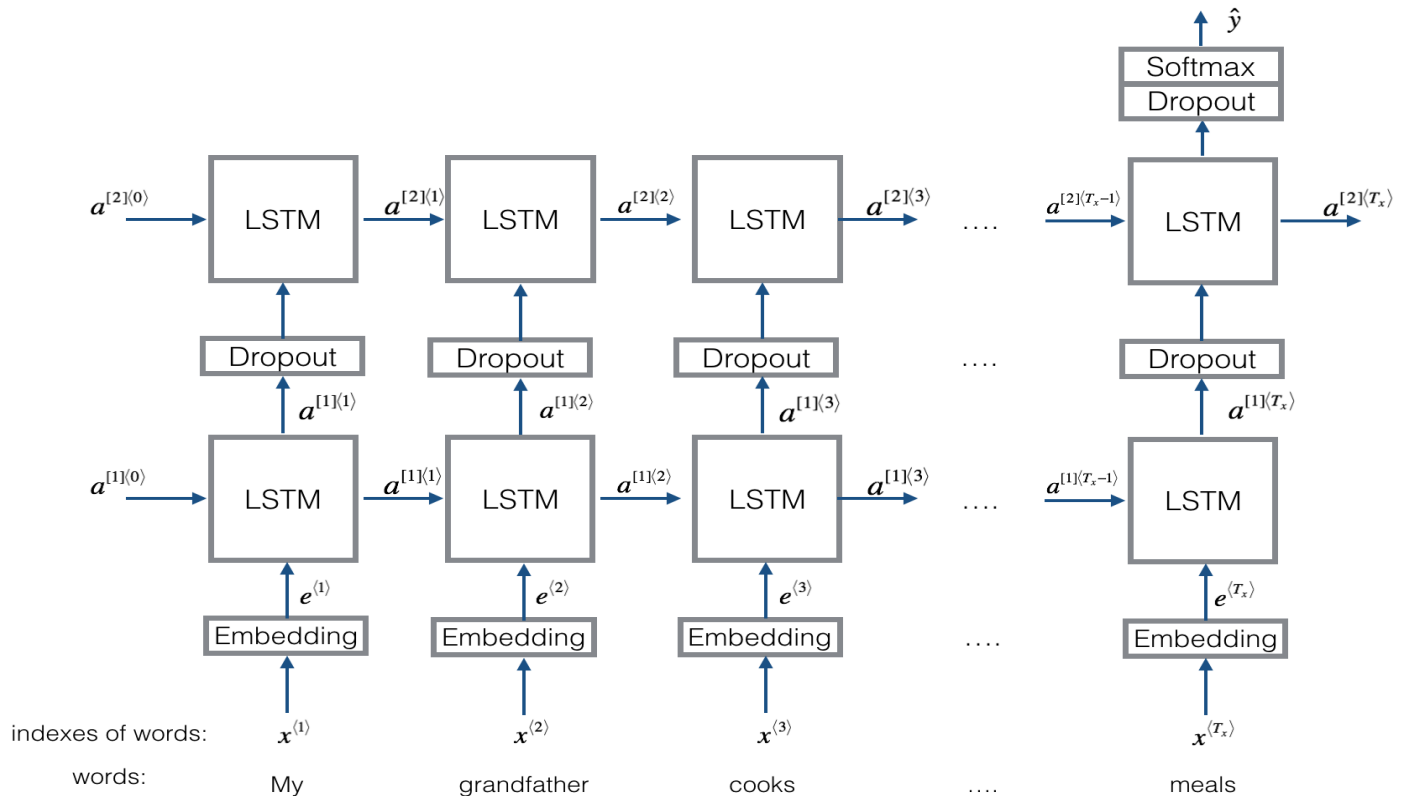


Figure 3: Emojiifier-v2. A 2-layer LSTM sequence classifier.

Exercise: Implement `Emojiify_V2()`, which builds a Keras graph of the architecture shown in Figure 3.

- The model takes as input an array of sentences of shape $(m, \text{max_len},)$ defined by `input_shape`.
- The model outputs a softmax probability vector of shape $(m, C = 5)$.
- You may need to use the following Keras layers:
 - `Input()` (<https://keras.io/layers/core/#input>)
 - Set the shape and dtype parameters.
 - The inputs are integers, so you can specify the data type as a string, 'int32'.
 - `LSTM()` (<https://keras.io/layers/recurrent/#lstm>)
 - Set the units and return_sequences parameters.
 - `Dropout()` (<https://keras.io/layers/core/#dropout>)
 - Set the rate parameter.
 - `Dense()` (<https://keras.io/layers/core/#dense>)
 - Set the units,
 - Note that `Dense()` has an activation parameter. For the purposes of passing the autograder, please do not set the activation within `Dense()`. Use the separate Activation layer to do so.
 - `Activation()` (<https://keras.io/activations/>).

- You can pass in the activation of your choice as a lowercase string.
- Model (<https://keras.io/models/model/>) Set inputs and outputs.

Additional Hints

- Remember that these Keras layers return an object, and you will feed in the outputs of the previous layer as the input arguments to that object. The returned object can be created and called in the same line.

How to use Keras layers in two lines of code

```
dense_object = Dense(units = ...)
```

```
X = dense_object(inputs)
```

How to use Keras layers in one line of code

```
X = Dense(units = ...)(inputs)
```

- The `embedding_layer` that is returned by `pretrained_embedding_layer` is a layer object that can be called as a function, passing in a single argument (sentence indices).
- Here is some sample code in case you're stuck

```
raw_inputs = Input(shape=(maxLen,), dtype='int32')
preprocessed_inputs = ... # some pre-processing
X = LSTM(units = ..., return_sequences= ...)(processed_inputs)
X = Dropout(rate = ..., )(X)
...
X = Dense(units = ...)(X)
X = Activation(...)(X)
model = Model(inputs=..., outputs=...)
...
```

In [47]: # GRADED FUNCTION: Emojify_V2

```
def Emojify_V2(input_shape, word_to_vec_map, word_to_index):
    """
    Function creating the Emojify-v2 model's graph.

    Arguments:
    input_shape -- shape of the input, usually (max_len,)
    word_to_vec_map -- dictionary mapping every word in a vocabulary into its
    50-dimensional vector representation
    word_to_index -- dictionary mapping from words to their indices in the voc
    abulary (400,001 words)

    Returns:
    model -- a model instance in Keras
    """

    ### START CODE HERE ###
    # Define sentence_indices as the input of the graph.
    # It should be of shape input_shape and dtype 'int32' (as it contains indi
    ces, which are integers).
    sentence_indices = Input(shape=input_shape, dtype='int32')

    # Create the embedding layer pretrained with GloVe Vectors (~1 line)
    embedding_layer = pretrained_embedding_layer(word_to_vec_map, word_to_inde
    x)

    # Propagate sentence_indices through your embedding layer
    # (See additional hints in the instructions).
    embeddings = embedding_layer(sentence_indices)

    # Propagate the embeddings through an LSTM layer with 128-dimensional hidd
    en state
    # The returned output should be a batch of sequences.
    X = LSTM(128, return_sequences=True)(embeddings)
    # Add dropout with a probability of 0.5
    X = Dropout(rate=0.5)(X)
    # Propagate X trough another LSTM layer with 128-dimensional hidden state
    # The returned output should be a single hidden state, not a batch of sequ
    ences.
    X = LSTM(128, return_sequences=False)(X) #according to the figure
    # Add dropout with a probability of 0.5
    X = Dropout(rate=0.5)(X)
    # Propagate X through a Dense layer with 5 units
    X = Dense(units=5)(X)
    # Add a softmax activation
    X = Activation(activation='softmax')(X)

    # Create Model instance which converts sentence_indices into X.
    model = Model(inputs=sentence_indices, outputs=X)

    ### END CODE HERE ###

    return model
```

Run the following cell to create your model and check its summary. Because all sentences in the dataset are less than 10 words, we chose `max_len = 10`. You should see your architecture, it uses "20,223,927" parameters, of which 20,000,050 (the word embeddings) are non-trainable, and the remaining 223,877 are. Because our vocabulary size has 400,001 words (with valid indices from 0 to 400,000) there are $400,001 * 50 = 20,000,050$ non-trainable parameters.

```
In [48]: model = Emojify_V2((maxLen,), word_to_vec_map, word_to_index)
        model.summary()
```

Layer (type)	Output Shape	Param #
=====		
input_1 (InputLayer)	(None, 10)	0
<hr/>		
embedding_2 (Embedding)	(None, 10, 50)	20000050
<hr/>		
lstm_1 (LSTM)	(None, 10, 128)	91648
<hr/>		
dropout_1 (Dropout)	(None, 10, 128)	0
<hr/>		
lstm_2 (LSTM)	(None, 128)	131584
<hr/>		
dropout_2 (Dropout)	(None, 128)	0
<hr/>		
dense_1 (Dense)	(None, 5)	645
<hr/>		
activation_1 (Activation)	(None, 5)	0
=====		
Total params: 20,223,927		
Trainable params: 223,877		
Non-trainable params: 20,000,050		
<hr/>		

As usual, after creating your model in Keras, you need to compile it and define what loss, optimizer and metrics you are want to use. Compile your model using `categorical_crossentropy` loss, `adam` optimizer and `['accuracy']` metrics:

```
In [49]: model.compile(loss='categorical_crossentropy', optimizer='adam', metrics=['accuracy'])
```

It's time to train your model. Your Emojifier-V2 model takes as input an array of shape `(m, max_len)` and outputs probability vectors of shape `(m, number of classes)`. We thus have to convert `X_train` (array of sentences as strings) to `X_train_indices` (array of sentences as list of word indices), and `Y_train` (labels as indices) to `Y_train_oh` (labels as one-hot vectors).

```
In [57]: X_train_indices = sentences_to_indices(X_train, word_to_index, maxLen)
        Y_train_oh = convert_to_one_hot(Y_train, C = 5)
```

Fit the Keras model on `X_train_indices` and `Y_train_oh`. We will use `epochs = 50` and `batch_size = 32`.

```
In [51]: model.fit(X_train_indices, Y_train_oh, epochs = 50, batch_size = 32, shuffle=True)
```



```
Epoch 1/50
132/132 [=====] - 0s - loss: 1.6105 - acc: 0.1667
Epoch 2/50
132/132 [=====] - 0s - loss: 1.5378 - acc: 0.3106
Epoch 3/50
132/132 [=====] - 0s - loss: 1.5060 - acc: 0.3030
Epoch 4/50
132/132 [=====] - 0s - loss: 1.4453 - acc: 0.3561
Epoch 5/50
132/132 [=====] - 0s - loss: 1.3588 - acc: 0.4167
Epoch 6/50
132/132 [=====] - 0s - loss: 1.2435 - acc: 0.5303
Epoch 7/50
132/132 [=====] - 0s - loss: 1.1799 - acc: 0.4394
Epoch 8/50
132/132 [=====] - 0s - loss: 1.0567 - acc: 0.5758
Epoch 9/50
132/132 [=====] - 0s - loss: 0.8761 - acc: 0.6970
Epoch 10/50
132/132 [=====] - 0s - loss: 0.8193 - acc: 0.7121
Epoch 11/50
132/132 [=====] - 0s - loss: 0.6981 - acc: 0.7500
Epoch 12/50
132/132 [=====] - 0s - loss: 0.5976 - acc: 0.8030
Epoch 13/50
132/132 [=====] - 0s - loss: 0.4894 - acc: 0.8258
Epoch 14/50
132/132 [=====] - 0s - loss: 0.5151 - acc: 0.8106
Epoch 15/50
132/132 [=====] - 0s - loss: 0.4852 - acc: 0.8106
Epoch 16/50
132/132 [=====] - 0s - loss: 0.3543 - acc: 0.8561
Epoch 17/50
132/132 [=====] - 0s - loss: 0.3960 - acc: 0.8636
Epoch 18/50
132/132 [=====] - 0s - loss: 0.6219 - acc: 0.8333
Epoch 19/50
132/132 [=====] - 0s - loss: 0.5163 - acc: 0.8182
Epoch 20/50
132/132 [=====] - 0s - loss: 0.3940 - acc: 0.8485
Epoch 21/50
132/132 [=====] - 0s - loss: 0.4647 - acc: 0.8106
Epoch 22/50
132/132 [=====] - 0s - loss: 0.3993 - acc: 0.8561
Epoch 23/50
132/132 [=====] - 0s - loss: 0.3735 - acc: 0.8409
Epoch 24/50
132/132 [=====] - 0s - loss: 0.3112 - acc: 0.9015
Epoch 25/50
132/132 [=====] - 0s - loss: 0.3435 - acc: 0.8864
Epoch 26/50
132/132 [=====] - 0s - loss: 0.2496 - acc: 0.9242
Epoch 27/50
132/132 [=====] - 0s - loss: 0.3098 - acc: 0.8788
Epoch 28/50
132/132 [=====] - 0s - loss: 0.2490 - acc: 0.9167
Epoch 29/50
```

```

132/132 [=====] - 0s - loss: 0.3731 - acc: 0.8636
Epoch 30/50
132/132 [=====] - 0s - loss: 0.2560 - acc: 0.9167
Epoch 31/50
132/132 [=====] - 0s - loss: 0.2879 - acc: 0.8864
Epoch 32/50
132/132 [=====] - 0s - loss: 0.1931 - acc: 0.9318
Epoch 33/50
132/132 [=====] - 0s - loss: 0.2042 - acc: 0.9545
Epoch 34/50
132/132 [=====] - 0s - loss: 0.1565 - acc: 0.9621
Epoch 35/50
132/132 [=====] - 0s - loss: 0.1663 - acc: 0.9621
Epoch 36/50
132/132 [=====] - 0s - loss: 0.1922 - acc: 0.9394
Epoch 37/50
132/132 [=====] - 0s - loss: 0.2515 - acc: 0.9167
Epoch 38/50
132/132 [=====] - 0s - loss: 0.2582 - acc: 0.9167
Epoch 39/50
132/132 [=====] - 0s - loss: 0.1524 - acc: 0.9470
Epoch 40/50
132/132 [=====] - 0s - loss: 0.1918 - acc: 0.9318
Epoch 41/50
132/132 [=====] - 0s - loss: 0.0913 - acc: 0.9773
Epoch 42/50
132/132 [=====] - 0s - loss: 0.1201 - acc: 0.9545
Epoch 43/50
132/132 [=====] - 0s - loss: 0.0948 - acc: 0.9621
Epoch 44/50
132/132 [=====] - 0s - loss: 0.0584 - acc: 0.9924
Epoch 45/50
132/132 [=====] - 0s - loss: 0.0711 - acc: 0.9848
Epoch 46/50
132/132 [=====] - 0s - loss: 0.0663 - acc: 0.9924
Epoch 47/50
132/132 [=====] - 0s - loss: 0.0581 - acc: 0.9924
Epoch 48/50
132/132 [=====] - 0s - loss: 0.0756 - acc: 0.9697
Epoch 49/50
132/132 [=====] - 0s - loss: 0.0468 - acc: 0.9924
Epoch 50/50
132/132 [=====] - 0s - loss: 0.0417 - acc: 0.9848

```

Out[51]: <keras.callbacks.History at 0x7f52d5097b00>

Your model should perform around **90% to 100% accuracy** on the training set. The exact accuracy you get may be a little different. Run the following cell to evaluate your model on the test set.

```
In [52]: X_test_indices = sentences_to_indices(X_test, word_to_index, max_len = maxLen)
Y_test_oh = convert_to_one_hot(Y_test, C = 5)
loss, acc = model.evaluate(X_test_indices, Y_test_oh)
print()
print("Test accuracy = ", acc)

32/56 [=====>.....] - ETA: 0s
Test accuracy = 0.928571428571
```

You should get a test accuracy between 80% and 95%. Run the cell below to see the mislabelled examples.

```
In [53]: # This code allows you to see the mislabelled examples
C = 5
y_test_oh = np.eye(C)[Y_test.reshape(-1)]
X_test_indices = sentences_to_indices(X_test, word_to_index, maxLen)
pred = model.predict(X_test_indices)
for i in range(len(X_test)):
    x = X_test_indices
    num = np.argmax(pred[i])
    if(num != Y_test[i]):
        print('Expected emoji:' + label_to_emoji(Y_test[i]) + ' prediction: ' +
X_test[i] + label_to_emoji(num).strip())

Expected emoji:😞 prediction: work is hard 😊
Expected emoji:😞 prediction: This girl is messing with me ❤️
Expected emoji:❤️ prediction: I love taking breaks 😞
Expected emoji:😞 prediction: you brighten my day ❤️
```

Now you can try it on your own example. Write your own sentence below.

```
In [54]: # Change the sentence below to see your prediction. Make sure all the words are in the Glove embeddings.
x_test = np.array(['not feeling happy'])
X_test_indices = sentences_to_indices(x_test, word_to_index, maxLen)
print(x_test[0] + ' ' + label_to_emoji(np.argmax(model.predict(X_test_indices))))

not feeling happy 😞
```

LSTM version accounts for word order

- Previously, Emojify-V1 model did not correctly label "not feeling happy," but our implementation of Emojify-V2 got it right.
 - (Keras' outputs are slightly random each time, so you may not have obtained the same result.)
- The current model still isn't very robust at understanding negation (such as "not happy")
 - This is because the training set is small and doesn't have a lot of examples of negation.
 - But if the training set were larger, the LSTM model would be much better than the Emojify-V1 model at understanding such complex sentences.

Congratulations!

You have completed this notebook! ❤️❤️❤️

What you should remember

- If you have an NLP task where the training set is small, using word embeddings can help your algorithm significantly.
- Word embeddings allow your model to work on words in the test set that may not even appear in the training set.
- Training sequence models in Keras (and in most other deep learning frameworks) requires a few important details:
 - To use mini-batches, the sequences need to be **padded** so that all the examples in a mini-batch have the **same length**.
 - An `Embedding()` layer can be initialized with pretrained values.
 - These values can be either fixed or trained further on your dataset.
 - If however your labeled dataset is small, it's usually not worth trying to train a large pre-trained set of embeddings.
 - `LSTM()` has a flag called `return_sequences` to decide if you would like to return every hidden states or only the last one.
 - You can use `Dropout()` right after `LSTM()` to regularize your network.

Input sentences:

"Congratulations on finishing this assignment and building an Emojifier."

"We hope you're happy with what you've accomplished in this notebook!"

Output emojis:



Acknowledgments

Thanks to Alison Darcy and the Woebot team for their advice on the creation of this assignment.

- Woebot is a chatbot friend that is ready to speak with you 24/7.
- Part of Woebot's technology uses word embeddings to understand the emotions of what you say.
- You can chat with Woebot by going to <http://woebot.io> (<http://woebot.io>).

