

Logistic Regression with a Neural Network mindset

Welcome to your first (required) programming assignment! You will build a logistic regression classifier to recognize cats. This assignment will step you through how to do this with a Neural Network mindset, and so will also hone your intuitions about deep learning.

Instructions:

- Do not use loops (for/while) in your code, unless the instructions explicitly ask you to do so.

You will learn to:

- Build the general architecture of a learning algorithm, including:
 - Initializing parameters
 - Calculating the cost function and its gradient
 - Using an optimization algorithm (gradient descent)
- Gather all three functions above into a main model function, in the right order.

Updates

This notebook has been updated over the past few months. The prior version was named "v5", and the current version is now named '6a'

If you were working on a previous version:

- You can find your prior work by looking in the file directory for the older files (named by version name).
- To view the file directory, click on the "Coursera" icon in the top left corner of this notebook.
- Please copy your work from the older versions to the new version, in order to submit your work for grading.

List of Updates

- Forward propagation formula, indexing now starts at 1 instead of 0.
- Optimization function comment now says "print cost every 100 training iterations" instead of "examples".
- Fixed grammar in the comments.
- `Y_prediction_test` variable name is used consistently.
- Plot's axis label now says "iterations (hundred)" instead of "iterations".
- When testing the model, the test image is normalized by dividing by 255.

1 - Packages

First, let's run the cell below to import all the packages that you will need during this assignment.

- [numpy](http://www.numpy.org) (www.numpy.org) is the fundamental package for scientific computing with Python.
- [h5py](http://www.h5py.org) (<http://www.h5py.org>) is a common package to interact with a dataset that is stored on an H5 file.
- [matplotlib](http://matplotlib.org) (<http://matplotlib.org>) is a famous library to plot graphs in Python.
- [PIL](http://www.pythonware.com/products/pil/) (<http://www.pythonware.com/products/pil/>) and [scipy](https://www.scipy.org/) (<https://www.scipy.org/>) are used here to test your model with your own picture at the end.

```
In [29]: import numpy as np
import matplotlib.pyplot as plt
import h5py
import scipy
from PIL import Image
from scipy import ndimage
from lr_utils import load_dataset

%matplotlib inline
```

2 - Overview of the Problem set

Problem Statement: You are given a dataset ("data.h5") containing:

- a training set of `m_train` images labeled as cat (`y=1`) or non-cat (`y=0`)
- a test set of `m_test` images labeled as cat or non-cat
- each image is of shape (`num_px, num_px, 3`) where 3 is for the 3 channels (RGB). Thus, each image is square (`height = num_px`) and (`width = num_px`).

You will build a simple image-recognition algorithm that can correctly classify pictures as cat or non-cat.

Let's get more familiar with the dataset. Load the data by running the following code.

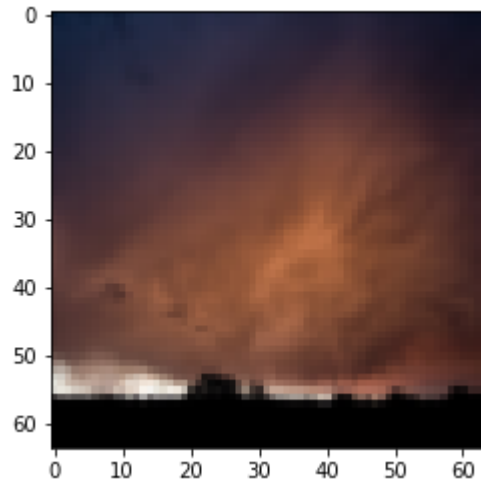
```
In [30]: # Loading the data (cat/non-cat)
train_set_x_orig, train_set_y, test_set_x_orig, test_set_y, classes = load_dataset()
```

We added "`_orig`" at the end of image datasets (train and test) because we are going to preprocess them. After preprocessing, we will end up with `train_set_x` and `test_set_x` (the labels `train_set_y` and `test_set_y` don't need any preprocessing).

Each line of your `train_set_x_orig` and `test_set_x_orig` is an array representing an image. You can visualize an example by running the following code. Feel free also to change the index value and re-run to see other images.

```
In [31]: # Example of a picture
index = 208
plt.imshow(train_set_x_orig[index])
print ("y = " + str(train_set_y[:, index]) + ", it's a '" + classes[np.squeeze
(train_set_y[:, index])).decode("utf-8") + "' picture.")
```

y = [0], it's a 'non-cat' picture.



Many software bugs in deep learning come from having matrix/vector dimensions that don't fit. If you can keep your matrix/vector dimensions straight you will go a long way toward eliminating many bugs.

Exercise: Find the values for:

- m_train (number of training examples)
- m_test (number of test examples)
- num_px (= height = width of a training image)

Remember that train_set_x_orig is a numpy-array of shape (m_train, num_px, num_px, 3). For instance, you can access m_train by writing train_set_x_orig.shape[0].

```
In [32]: ### START CODE HERE ### (~ 3 lines of code)
print(train_set_x_orig.shape)      #(290, 64, 64, 3) -> (no of training images,
height, width, 3 channels of rgb)

m_train = train_set_x_orig.shape[0]
m_test = test_set_x_orig.shape[0]
num_px = train_set_x_orig.shape[2]  #no of pixels height or width
### END CODE HERE ###

print ("Number of training examples: m_train = " + str(m_train))
print ("Number of testing examples: m_test = " + str(m_test))
print ("Height/Width of each image: num_px = " + str(num_px))
print ("Each image is of size: (" + str(num_px) + ", " + str(num_px) + ", 3)")
print ("train_set_x shape: " + str(train_set_x_orig.shape))
print ("train_set_y shape: " + str(train_set_y.shape))
print ("test_set_x shape: " + str(test_set_x_orig.shape))
print ("test_set_y shape: " + str(test_set_y.shape))

(209, 64, 64, 3)
Number of training examples: m_train = 209
Number of testing examples: m_test = 50
Height/Width of each image: num_px = 64
Each image is of size: (64, 64, 3)
train_set_x shape: (209, 64, 64, 3)
train_set_y shape: (1, 209)
test_set_x shape: (50, 64, 64, 3)
test_set_y shape: (1, 50)
```

Expected Output for m_train, m_test and num_px:

m_train	209
m_test	50
num_px	64

For convenience, you should now reshape images of shape (num_px, num_px, 3) in a numpy-array of shape (num_px * num_px * 3, 1). After this, our training (and test) dataset is a numpy-array where each column represents a flattened image. There should be m_train (respectively m_test) columns.

Exercise: Reshape the training and test data sets so that images of size (num_px, num_px, 3) are flattened into single vectors of shape (num_px * num_px * 3, 1).

A trick when you want to flatten a matrix X of shape (a,b,c,d) to a matrix X_flatten of shape (b*c*d, a) is to use:

```
X_flatten = X.reshape(X.shape[0], -1).T      # X.T is the transpose of X
```

```
In [33]: # Reshape the training and test examples
#flattening our datasets (remember we're now working with the entire dataset), into a column vector, in other words
#stacking each image on top of each other

### START CODE HERE ### (= 2 lines of code)
#all the train, test images are flattened into a single column matrix, this way its easier to process
train_set_x_flatten = train_set_x_orig.reshape(train_set_x_orig.shape[0], -1).T #flattening trick
test_set_x_flatten = test_set_x_orig.reshape(test_set_x_orig.shape[0], -1).T
### END CODE HERE ###

print ("train_set_x_flatten shape: " + str(train_set_x_flatten.shape))
print ("train_set_y shape: " + str(train_set_y.shape))
print ("test_set_x_flatten shape: " + str(test_set_x_flatten.shape))
print ("test_set_y shape: " + str(test_set_y.shape))
print ("sanity check after reshaping: " + str(train_set_x_flatten[0:5,0]))

train_set_x_flatten shape: (12288, 209)
train_set_y shape: (1, 209)
test_set_x_flatten shape: (12288, 50)
test_set_y shape: (1, 50)
sanity check after reshaping: [17 31 56 22 33]
```

Expected Output:

train_set_x_flatten shape	(12288, 209)
train_set_y shape	(1, 209)
test_set_x_flatten shape	(12288, 50)
test_set_y shape	(1, 50)
sanity check after reshaping	[17 31 56 22 33]

To represent color images, the red, green and blue channels (RGB) must be specified for each pixel, and so the pixel value is actually a vector of three numbers ranging from 0 to 255.

One common preprocessing step in machine learning is to center and standardize your dataset, meaning that you subtract the mean of the whole numpy array from each example, and then divide each example by the standard deviation of the whole numpy array. But for picture datasets, it is simpler and more convenient and works almost as well to just divide every row of the dataset by 255 (the maximum value of a pixel channel).

Let's standardize our dataset.

```
In [34]: train_set_x = train_set_x_flatten/255.  
        test_set_x = test_set_x_flatten/255.
```

What you need to remember:

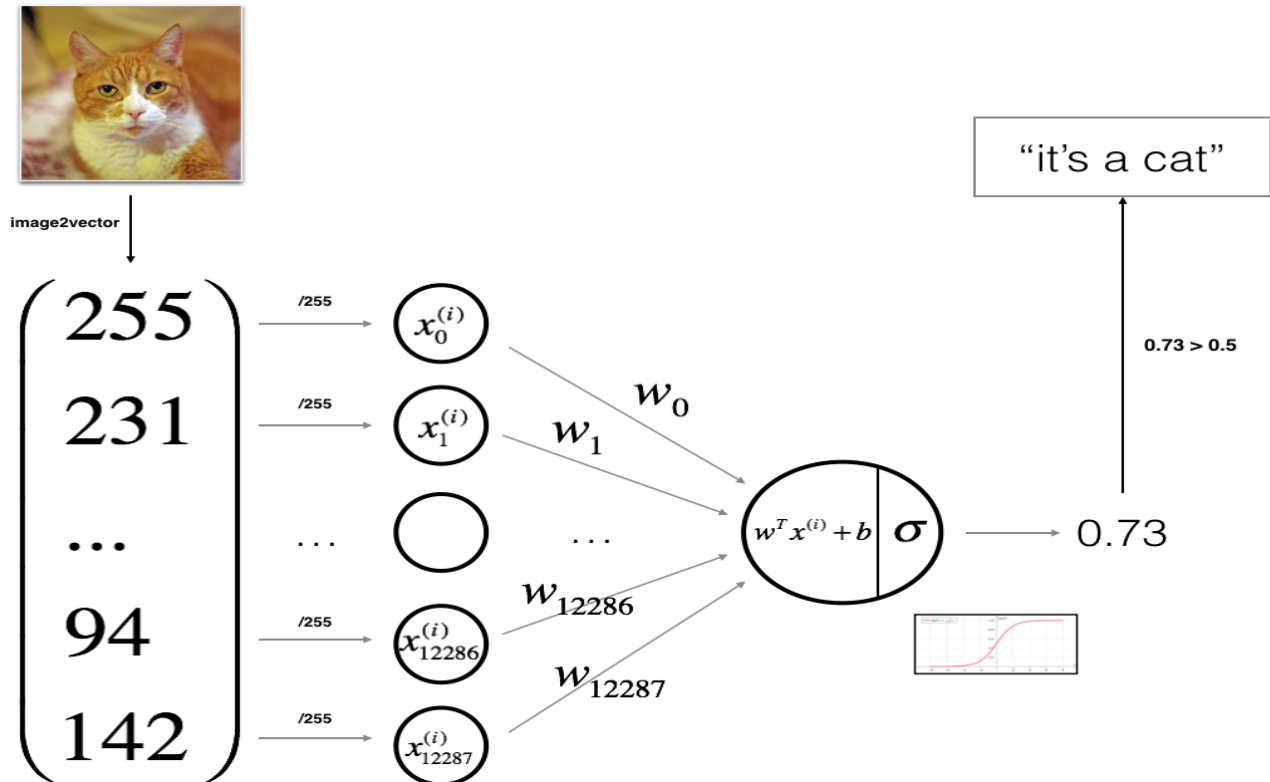
Common steps for pre-processing a new dataset are:

- Figure out the dimensions and shapes of the problem (m_{train} , m_{test} , num_px , ...)
- Reshape the datasets such that each example is now a vector of size ($\text{num_px} * \text{num_px} * 3, 1$)
- "Standardize" the data

3 - General Architecture of the learning algorithm

It's time to design a simple algorithm to distinguish cat images from non-cat images.

You will build a Logistic Regression, using a Neural Network mindset. The following Figure explains why **Logistic Regression is actually a very simple Neural Network!**



Mathematical expression of the algorithm:

For one example $x^{(i)}$:

$$z^{(i)} = w^T x^{(i)} + b \quad (1)$$

$$\hat{y}^{(i)} = a^{(i)} = \text{sigmoid}(z^{(i)}) \quad (2)$$

$$\mathcal{L}(a^{(i)}, y^{(i)}) = -y^{(i)} \log(a^{(i)}) - (1 - y^{(i)}) \log(1 - a^{(i)}) \quad (3)$$

The cost is then computed by summing over all training examples:

$$J = \frac{1}{m} \sum_{i=1}^m \mathcal{L}(a^{(i)}, y^{(i)}) \quad (6)$$

Key steps: In this exercise, you will carry out the following steps:

- Initialize the parameters of the model
- Learn the parameters for the model by minimizing the cost
- Use the learned parameters to make predictions (on the test set)
- Analyse the results and conclude

4 - Building the parts of our algorithm

The main steps for building a Neural Network are:

1. Define the model structure (such as number of input features)
2. Initialize the model's parameters
3. Loop:
 - Calculate current loss (forward propagation)
 - Calculate current gradient (backward propagation)
 - Update parameters (gradient descent)

You often build 1-3 separately and integrate them into one function we call `model()`.

4.1 - Helper functions

Exercise: Using your code from "Python Basics", implement `sigmoid()`. As you've seen in the figure above, you need to compute $\text{sigmoid}(w^T x + b) = \frac{1}{1 + e^{-(w^T x + b)}}$ to make predictions. Use `np.exp()`.

```
In [35]: # GRADED FUNCTION: sigmoid
#the activation function
def sigmoid(z):
    """
    Compute the sigmoid of z

    Arguments:
    z -- A scalar or numpy array of any size.

    Return:
    s -- sigmoid(z)
    """

    ### START CODE HERE ### (~ 1 line of code)
    s = 1/(1+np.exp(-z)) #the sigmoid function -> 1/1 + e^-z
    ### END CODE HERE ###

    return s
```

```
In [36]: print ("sigmoid([0, 2]) = " + str(sigmoid(np.array([0,2]))))

sigmoid([0, 2]) = [ 0.5          0.88079708]
```

Expected Output:

sigmoid([0, 2])	[0.5 0.88079708]
---------------------	-------------------

4.2 - Initializing parameters

Exercise: Implement parameter initialization in the cell below. You have to initialize w as a vector of zeros. If you don't know what numpy function to use, look up `np.zeros()` in the Numpy library's documentation.

```
In [37]: # GRADED FUNCTION: initialize_with_zeros
#initializes values for w and b, matrix sizes are as follows
#w -> (dim, dim-1) b -> (dim, 1)
def initialize_with_zeros(dim):
    """
    This function creates a vector of zeros of shape (dim, 1) for w and initializes b to 0.

    Argument:
    dim -- size of the w vector we want (or number of parameters in this case)

    Returns:
    w -- initialized vector of shape (dim, 1)
    b -- initialized scalar (corresponds to the bias)
    """

    ### START CODE HERE ### (= 1 line of code)
    w = np.zeros((dim, dim-1))
    b = 0 #initializing b = 0 here, but from next week it turns into a vector (dim, 1)
    ### END CODE HERE ###

    assert(w.shape == (dim, 1))
    assert(isinstance(b, float) or isinstance(b, int))

    return w, b
```

```
In [38]: dim = 2
w, b = initialize_with_zeros(dim)
print ("w = " + str(w))
print ("b = " + str(b))

w = [[ 0.]
      [ 0.]]
b = 0
```

Expected Output:

** w **	[[0.] [0.]]
** b **	0

For image inputs, w will be of shape $(\text{num_px} \times \text{num_px} \times 3, 1)$.

4.3 - Forward and Backward propagation

Now that your parameters are initialized, you can do the "forward" and "backward" propagation steps for learning the parameters.

Exercise: Implement a function `propagate()` that computes the cost function and its gradient.

Hints:

Forward Propagation:

- You get X
- You compute $A = \sigma(w^T X + b) = (a^{(1)}, a^{(2)}, \dots, a^{(m-1)}, a^{(m)})$
- You calculate the cost function: $J = -\frac{1}{m} \sum_{i=1}^m y^{(i)} \log(a^{(i)}) + (1 - y^{(i)}) \log(1 - a^{(i)})$

Here are the two formulas you will be using:

$$\frac{\partial J}{\partial w} = \frac{1}{m} X(A - Y)^T \quad (7)$$

$$\frac{\partial J}{\partial b} = \frac{1}{m} \sum_{i=1}^m (a^{(i)} - y^{(i)}) \quad (8)$$

```

In [39]: # GRADED FUNCTION: propagate
#forward and backward prop function, computes the cost, the corresponding yhat
and the derivatives for each iteration
#performed in the optimize function
def propagate(w, b, X, Y):
    """
    Implement the cost function and its gradient for the propagation explained
    above

    Arguments:
    w -- weights, a numpy array of size (num_px * num_px * 3, 1)
    b -- bias, a scalar
    X -- data of size (num_px * num_px * 3, number of examples)
    Y -- true "label" vector (containing 0 if non-cat, 1 if cat) of size (1, n
    umber of examples)

    Return:
    cost -- negative log-likelihood cost for logistic regression
    dw -- gradient of the loss with respect to w, thus same shape as w
    db -- gradient of the loss with respect to b, thus same shape as b

    Tips:
    - Write your code step by step for the propagation. np.log(), np.dot()
    """

    m = X.shape[1]

    # FORWARD PROPAGATION (FROM X TO COST)
    ### START CODE HERE ### (≈ 2 lines of code)
    #calculation for entire training, via vectorization, yhat (A) and cost for
each iteration
    A = sigmoid((np.dot(w.T, X)) + b)      #A = sigmoid(Z)                # compute activation
    cost = -1/m * np.sum((Y * np.log(A) + (1-Y) * np.log(1-A)))          # compute cost
    ### END CODE HERE ###

    # BACKWARD PROPAGATION (TO FIND GRAD)
    ### START CODE HERE ### (≈ 2 lines of code)
    dw = 1/m * np.dot(X, (A-Y).T)        #1/m * dZ.T . X
    db = 1/m * np.sum(A-Y)                #1/m * summation of dZ
    ### END CODE HERE ###

    assert(dw.shape == w.shape)
    assert(db.dtype == float)
    cost = np.squeeze(cost)
    assert(cost.shape == ())

    grads = {"dw": dw,
              "db": db}

    return grads, cost

```

```
In [40]: w, b, X, Y = np.array([[1.],[2.]]), 2., np.array([[1.,2.,-1.],[3.,4.,-3.2]]),
np.array([[1,0,1]])
grads, cost = propagate(w, b, X, Y)
print ("dw = " + str(grads["dw"]))
print ("db = " + str(grads["db"]))
print ("cost = " + str(cost))
```

```
dw = [[ 0.99845601]
 [ 2.39507239]]
db = 0.00145557813678
cost = 5.80154531939
```

Expected Output:

** dw **	[[0.99845601] [2.39507239]]
** db **	0.00145557813678
** cost **	5.801545319394553

4.4 - Optimization

- You have initialized your parameters.
- You are also able to compute a cost function and its gradient.
- Now, you want to update the parameters using gradient descent.

Exercise: Write down the optimization function. The goal is to learn w and b by minimizing the cost function J . For a parameter θ , the update rule is $\theta = \theta - \alpha d\theta$, where α is the learning rate.

```
In [41]: # GRADED FUNCTION: optimize
        ##w is basically a random initial point we pick on the cost graph, b is a simple bias value
        #this function iterates thru a given number of times, where each iteration performs one descent in the cost graph, updating
        #the w and b values respectively depending on their derivatives
        #so basically this means, the greater the number of iterations the more likely we'll reach the global optimum

        #upon each iteration we obtain a different value for b and w cuz of the updates performed on it, so what happens is that
        #the value of w keeps decreasing yea (cuz it gets deducted each iteration), decreasing means that the random point that we
        #pick moves down the curve (descends from our initial start position), as mentioned before, our goal is to find w and b
        #values that result in the lowest possible cost function, and that is found by moving closer and closer towards the global
        #optimum of the curve (where the curve has a gradient of 0)
        def optimize(w, b, X, Y, num_iterations, learning_rate, print_cost = False):
            """
            This function optimizes w and b by running a gradient descent algorithm
```

Arguments:

w -- weights, a numpy array of size (num_px * num_px * 3, 1)
b -- bias, a scalar
X -- data of shape (num_px * num_px * 3, number of examples)
Y -- true "label" vector (containing 0 if non-cat, 1 if cat), of shape (1, number of examples)
num_iterations -- number of iterations of the optimization loop
learning_rate -- learning rate of the gradient descent update rule
print_cost -- True to print the loss every 100 steps

Returns:

params -- dictionary containing the weights *w* and bias *b*
grads -- dictionary containing the gradients of the weights and bias with respect to the cost function
costs -- list of all the costs computed during the optimization, this will be used to plot the learning curve.

Tips:

You basically need to write down two steps and iterate through them:

1) Calculate the cost and the gradient for the current parameters. Use `propagate()`.

2) Update the parameters using gradient descent rule for *w* and *b*.

"""

`costs = []`

`for i in range(num_iterations):`

Cost and gradient calculation (\approx 1-4 lines of code)

START CODE HERE

`grads, cost = propagate(w, b, X, Y)`

END CODE HERE

Retrieve derivatives from grads

`dw = grads["dw"]`

`db = grads["db"]`

update rule (\approx 2 lines of code)

START CODE HERE

`w = w - (learning_rate * dw)`

`b = b - (learning_rate * db) #param = param - (alpha(learning rate) * param)`

END CODE HERE

Record the costs

`if i % 100 == 0:`

`costs.append(cost)`

Print the cost every 100 training iterations

`if print_cost and i % 100 == 0:`

`print ("Cost after iteration %i: %f" % (i, cost))`

`params = {"w": w,
 "b": b}`

`grads = {"dw": dw,`

```
"db": db}
```

```
return params, grads, costs
```

```
In [42]: params, grads, costs = optimize(w, b, X, Y, num_iterations= 100, learning_rate
      = 0.009, print_cost = False)
```

```
print ("w = " + str(params["w"]))
print ("b = " + str(params["b"]))
print ("dw = " + str(grads["dw"]))
print ("db = " + str(grads["db"]))
```

```
w = [[ 0.19033591]
      [ 0.12259159]]
b = 1.92535983008
dw = [[ 0.67752042]
       [ 1.41625495]]
db = 0.219194504541
```

Expected Output:

w	[[0.19033591] [0.12259159]]
b	1.92535983008
dw	[[0.67752042] [1.41625495]]
db	0.219194504541

Exercise: The previous function will output the learned w and b . We are able to use w and b to predict the labels for a dataset X . Implement the `predict()` function. There are two steps to computing predictions:

1. Calculate $\hat{Y} = A = \sigma(w^T X + b)$
2. Convert the entries of A into 0 (if activation ≤ 0.5) or 1 (if activation > 0.5), stores the predictions in a vector `Y_prediction`. If you wish, you can use an `if/else` statement in a `for` loop (though there is also a way to vectorize this).

```

In [43]: # GRADED FUNCTION: predict
#scales the prediction values, after calculating it ofc
def predict(w, b, X):
    '''
    Predict whether the label is 0 or 1 using Learned logistic regression parameters (w, b)

    Arguments:
    w -- weights, a numpy array of size (num_px * num_px * 3, 1)
    b -- bias, a scalar
    X -- data of size (num_px * num_px * 3, number of examples)

    Returns:
    Y_prediction -- a numpy array (vector) containing all predictions (0/1) for the examples in X
    '''

    m = X.shape[1]
    Y_prediction = np.zeros((1,m))
    w = w.reshape(X.shape[0], 1)

    # Compute vector "A" predicting the probabilities of a cat being present in the picture
    ### START CODE HERE ### (~ 1 line of code)
    #this A is the final output A
    A = sigmoid(np.dot(w.T, X) + b) #A has the yhat values for each image
    ### END CODE HERE ###

    for i in range(A.shape[1]):

        # Convert probabilities A[0,i] to actual predictions p[0,i]
        ### START CODE HERE ### (~ 4 lines of code)
        if A[0][i] >= 0.5:
            Y_prediction[0][i] = 1
        else:
            Y_prediction[0][i] = 0
        ### END CODE HERE ###

    assert(Y_prediction.shape == (1, m))

    return Y_prediction

```

```

In [44]: w = np.array([[0.1124579],[0.23106775]])
b = -0.3
X = np.array([[1.,-1.1,-3.2],[1.2,2.,0.1]])
print ("predictions = " + str(predict(w, b, X)))

predictions = [[ 1.  1.  0.]]

```

Expected Output:

predictions	[[1. 1. 0.]]
-----------------	-----------------

What to remember: You've implemented several functions that:

- Initialize (w, b)
- Optimize the loss iteratively to learn parameters (w, b) :
 - computing the cost and its gradient
 - updating the parameters using gradient descent
- Use the learned (w, b) to predict the labels for a given set of examples

5 - Merge all functions into a model

You will now see how the overall model is structured by putting together all the building blocks (functions implemented in the previous parts) together, in the right order.

Exercise: Implement the model function. Use the following notation:

- `Y_prediction_test` for your predictions on the test set
- `Y_prediction_train` for your predictions on the train set
- `w, costs, grads` for the outputs of `optimize()`

In [117]: *# GRADED FUNCTION: model*

```
def model(X_train, Y_train, X_test, Y_test, num_iterations = 2000, learning_rate = 0.5, print_cost = False):
    """
    Builds the logistic regression model by calling the function you've implemented previously

    Arguments:
    X_train -- training set represented by a numpy array of shape (num_px * num_px * 3, m_train)
    Y_train -- training labels represented by a numpy array (vector) of shape (1, m_train)
    X_test -- test set represented by a numpy array of shape (num_px * num_px * 3, m_test)
    Y_test -- test labels represented by a numpy array (vector) of shape (1, m_test)
    num_iterations -- hyperparameter representing the number of iterations to optimize the parameters
    learning_rate -- hyperparameter representing the learning rate used in the update rule of optimize()
    print_cost -- Set to true to print the cost every 100 iterations

    Returns:
    d -- dictionary containing information about the model.
    """

    ### START CODE HERE ###

    # initialize parameters with zeros (~ 1 line of code)
    w, b = initialize_with_zeros(X_train.reshape(X_train.shape[0], -1).T.shape[1])

    # Gradient descent (~ 1 line of code)
    parameters, grads, costs = optimize(w, b, X_train, Y_train, num_iterations, learning_rate, print_cost)

    # Retrieve parameters w and b from dictionary "parameters"
    w = parameters["w"]
    b = parameters["b"]

    # Predict test/train set examples (~ 2 lines of code)
    Y_prediction_test = predict(w, b, X_test)
    Y_prediction_train = predict(w, b, X_train)

    ### END CODE HERE ###

    # Print train/test Errors
    print("train accuracy: {} %".format(100 - np.mean(np.abs(Y_prediction_train - Y_train)) * 100))
    print("test accuracy: {} %".format(100 - np.mean(np.abs(Y_prediction_test - Y_test)) * 100))

    d = {"costs": costs,
        "Y_prediction_test": Y_prediction_test,
```

```

"Y_prediction_train" : Y_prediction_train,
"w" : w,
"b" : b,
"learning_rate" : learning_rate,
"num_iterations": num_iterations}

return d

```

Run the following cell to train your model.

```
In [118]: d = model(train_set_x, train_set_y, test_set_x, test_set_y, num_iterations = 2000, learning_rate = 0.005, print_cost = True)
```

```

Cost after iteration 0: 0.693147
Cost after iteration 100: 0.584508
Cost after iteration 200: 0.466949
Cost after iteration 300: 0.376007
Cost after iteration 400: 0.331463
Cost after iteration 500: 0.303273
Cost after iteration 600: 0.279880
Cost after iteration 700: 0.260042
Cost after iteration 800: 0.242941
Cost after iteration 900: 0.228004
Cost after iteration 1000: 0.214820
Cost after iteration 1100: 0.203078
Cost after iteration 1200: 0.192544
Cost after iteration 1300: 0.183033
Cost after iteration 1400: 0.174399
Cost after iteration 1500: 0.166521
Cost after iteration 1600: 0.159305
Cost after iteration 1700: 0.152667
Cost after iteration 1800: 0.146542
Cost after iteration 1900: 0.140872
train accuracy: 99.04306220095694 %
test accuracy: 70.0 %

```

Expected Output:

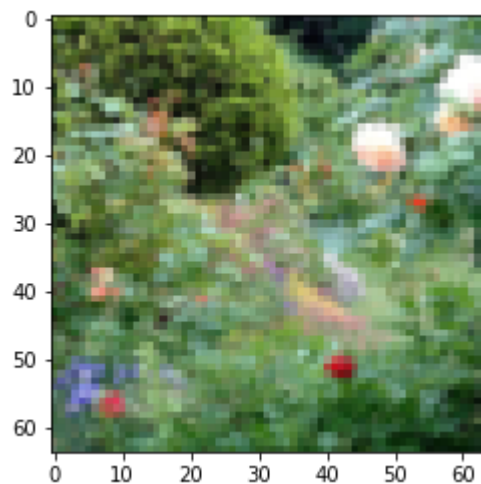
**Cost after iteration 0 **	0.693147
:	:
Train Accuracy	99.04306220095694 %
Test Accuracy	70.0 %

Comment: Training accuracy is close to 100%. This is a good sanity check: your model is working and has high enough capacity to fit the training data. Test accuracy is 68%. It is actually not bad for this simple model, given the small dataset we used and that logistic regression is a linear classifier. But no worries, you'll build an even better classifier next week!

Also, you see that the model is clearly overfitting the training data. Later in this specialization you will learn how to reduce overfitting, for example by using regularization. Using the code below (and changing the index variable) you can look at predictions on pictures of the test set.

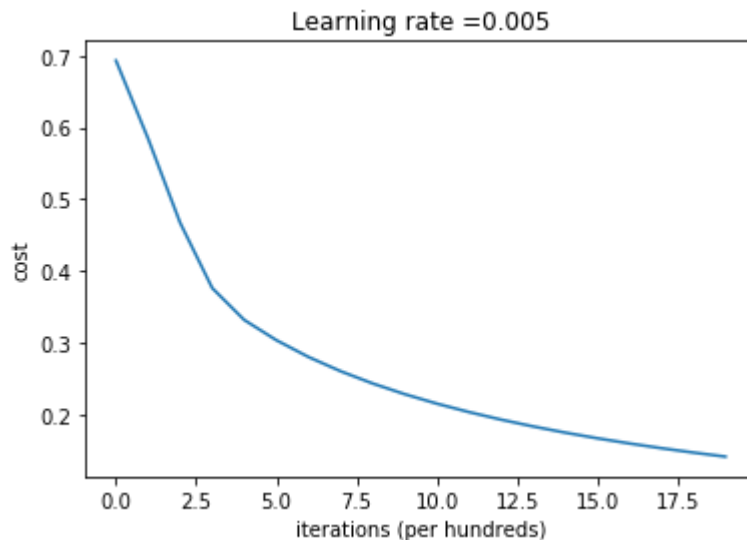
```
In [76]: # Example of a picture that was wrongly classified.  
index = 49  
plt.imshow(test_set_x[:,index].reshape((num_px, num_px, 3)))  
print ("y = " + str(test_set_y[0,index]) + ", you predicted that it is a \"" +  
classes[d["Y_prediction_test"][0,index]].decode("utf-8") + "\" picture.")
```

y = 0, you predicted that it is a "non-cat" picture.



Let's also plot the cost function and the gradients.

```
In [77]: # Plot Learning curve (with costs)
costs = np.squeeze(d['costs'])
plt.plot(costs)
plt.ylabel('cost')
plt.xlabel('iterations (per hundreds)')
plt.title("Learning rate =" + str(d["learning_rate"]))
plt.show()
```



Interpretation: You can see the cost decreasing. It shows that the parameters are being learned. However, you see that you could train the model even more on the training set. Try to increase the number of iterations in the cell above and rerun the cells. You might see that the training set accuracy goes up, but the test set accuracy goes down. This is called overfitting.

6 - Further analysis (optional/ungraded exercise)

Congratulations on building your first image classification model. Let's analyze it further, and examine possible choices for the learning rate α .

Choice of learning rate

Reminder: In order for Gradient Descent to work you must choose the learning rate wisely. The learning rate α determines how rapidly we update the parameters. If the learning rate is too large we may "overshoot" the optimal value. Similarly, if it is too small we will need too many iterations to converge to the best values. That's why it is crucial to use a well-tuned learning rate.

Let's compare the learning curve of our model with several choices of learning rates. Run the cell below. This should take about 1 minute. Feel free also to try different values than the three we have initialized the `learning_rates` variable to contain, and see what happens.

```

In [78]: learning_rates = [0.01, 0.001, 0.0001]
models = {}
for i in learning_rates:
    print ("learning rate is: " + str(i))
    models[str(i)] = model(train_set_x, train_set_y, test_set_x, test_set_y, n
um_iterations = 1500, learning_rate = i, print_cost = False)
    print ('\n' + "-----" +
'\n')

for i in learning_rates:
    plt.plot(np.squeeze(models[str(i)]["costs"]), label= str(models[str(i)]["l
earning_rate"]))

plt.ylabel('cost')
plt.xlabel('iterations (hundreds)')

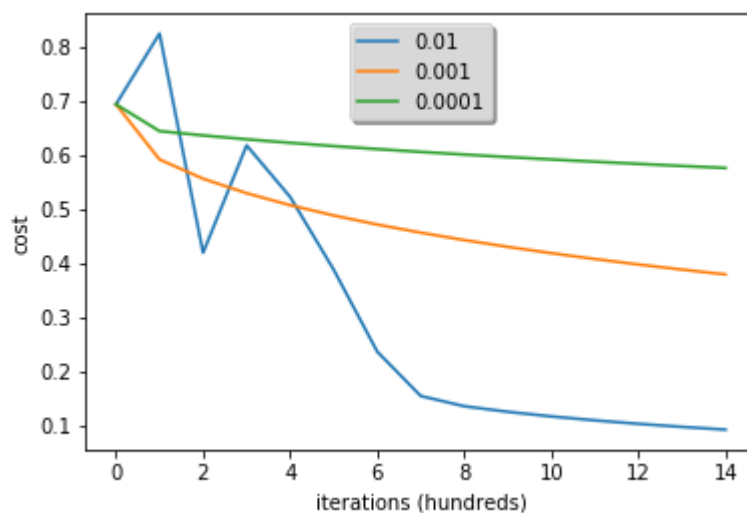
legend = plt.legend(loc='upper center', shadow=True)
frame = legend.get_frame()
frame.set_facecolor('0.90')
plt.show()

```

learning rate is: 0.01
train accuracy: 99.52153110047847 %
test accuracy: 68.0 %

learning rate is: 0.001
train accuracy: 88.99521531100478 %
test accuracy: 64.0 %

learning rate is: 0.0001
train accuracy: 68.42105263157895 %
test accuracy: 36.0 %



Interpretation:

- Different learning rates give different costs and thus different predictions results.
- If the learning rate is too large (0.01), the cost may oscillate up and down. It may even diverge (though in this example, using 0.01 still eventually ends up at a good value for the cost).
- A lower cost doesn't mean a better model. You have to check if there is possibly overfitting. It happens when the training accuracy is a lot higher than the test accuracy.
- In deep learning, we usually recommend that you:
 - Choose the learning rate that better minimizes the cost function.
 - If your model overfits, use other techniques to reduce overfitting. (We'll talk about this in later videos.)

7 - Test with your own image (optional/ungraded exercise)

Congratulations on finishing this assignment. You can use your own image and see the output of your model. To do that:

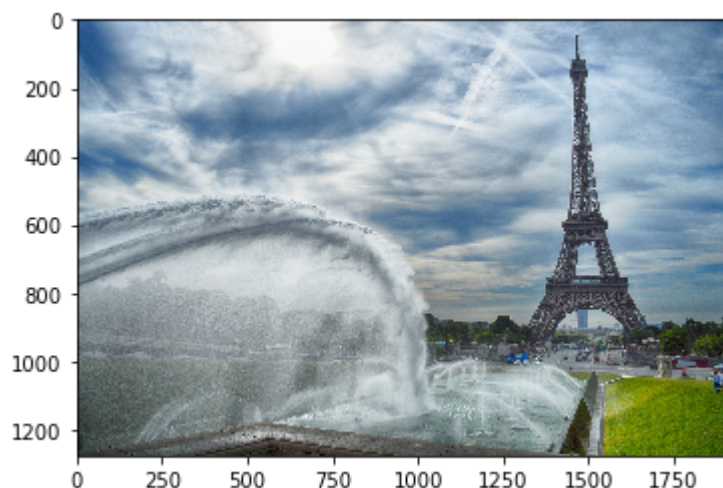
1. Click on "File" in the upper bar of this notebook, then click "Open" to go on your Coursera Hub.
2. Add your image to this Jupyter Notebook's directory, in the "images" folder
3. Change your image's name in the following code
4. Run the code and check if the algorithm is right (1 = cat, 0 = non-cat)!

```
In [79]: ## START CODE HERE ## (PUT YOUR IMAGE NAME)
my_image = "my_image.jpg" # change this to the name of your image file
## END CODE HERE ##

# We preprocess the image to fit your algorithm.
fname = "images/" + my_image
image = np.array(ndimage.imread(fname, flatten=False))
image = image/255.
my_image = scipy.misc.imresize(image, size=(num_px,num_px)).reshape((1, num_px
*num_px*3)).T
my_predicted_image = predict(d["w"], d["b"], my_image)

plt.imshow(image)
print("y = " + str(np.squeeze(my_predicted_image)) + ", your algorithm predict
s a \" + classes[int(np.squeeze(my_predicted_image)),].decode("utf-8") + "\"
picture.")
```

$y = 0.0$, your algorithm predicts a "non-cat" picture.



What to remember from this assignment:

1. Preprocessing the dataset is important.
2. You implemented each function separately: initialize(), propagate(), optimize(). Then you built a model().
3. Tuning the learning rate (which is an example of a "hyperparameter") can make a big difference to the algorithm. You will see more examples of this later in this course!

Finally, if you'd like, we invite you to try different things on this Notebook. Make sure you submit before trying anything. Once you submit, things you can play with include:

- Play with the learning rate and the number of iterations
- Try different initialization methods and compare the results
- Test other preprocessings (center the data, or divide each row by its standard deviation)

Bibliography:

- <http://www.wildml.com/2015/09/implementing-a-neural-network-from-scratch/>
(<http://www.wildml.com/2015/09/implementing-a-neural-network-from-scratch/>)
- <https://stats.stackexchange.com/questions/211436/why-do-we-normalize-images-by-subtracting-the-datasets-image-mean-and-not-the-c> (<https://stats.stackexchange.com/questions/211436/why-do-we-normalize-images-by-subtracting-the-datasets-image-mean-and-not-the-c>)

In []: