Министерство образования Республики Беларусь

Учреждение образования

“Белорусский государственный университет информатики и радиоэлектроники”

Факультет компьютерных систем и сетей

Специальность “Обработка больших объемов информации”

Лабораторная работа №4

**Нейронные сети**

Выполнил:

магистрант гр. 858641 Кальман В.А.

Проверил:

Стержанов М. В.

Минск 2019

**Данные.**

Набор данных ex4data1.mat (такой же, как в лабораторной работе №2) представляет собой файл формата \*.mat (т.е. сохраненного из Matlab). Набор содержит 5000 изображений 20x20 в оттенках серого. Каждый пиксель представляет собой значение яркости (вещественное число). Каждое изображение сохранено в виде вектора из 400 элементов. В результате загрузки набора данных должна быть получена матрица 5000x400. Далее расположены метки классов изображений от 1 до 9 (соответствуют цифрам от 1 до 9), а также 10 (соответствует цифре 0).

**1. Загрузите данные ex4data1.mat из файла.**

**from** **copy** **import** deepcopy

**import** **numpy** **as** **np**

**import** **matplotlib.pyplot** **as** **plt**

**from** **scipy** **import** io

**from** **sklearn.preprocessing** **import** OneHotEncoder

%matplotlib inline

ex4data1 = io.loadmat('Data/Lab 4/ex4data1.mat')

X, y = ex4data1['X'], ex4data1['y'].flatten()

**2. Загрузите веса нейронной сети из файла ex4weights.mat, который содержит две матрицы Θ(1) (25, 401) и Θ(2) (10, 26). Какова структура полученной нейронной сети?**

ex4weights = io.loadmat('Data/Lab 4/ex4weights.mat')

theta1, theta2 = ex4weights['Theta1'], ex4weights['Theta2']

Структура сети: двухслойная полносвязная сеть

**3. Реализуйте функцию прямого распространения с сигмоидом в качестве функции активации.**

def sigmoid(z):

return 1 / (1 + np.exp(-z))

def h0x(X, theta):

m = len(X)

ones = np.ones((m, 1))

a1 = np.hstack((ones, X))

a2 = sigmoid(np.dot(a1, theta[0].T))

a2 = np.hstack((ones, a2))

h = sigmoid(np.dot(a2, theta[1].T))

return h

**4. Вычислите процент правильных классификаций на обучающей выборке. Сравните полученный результат с логистической регрессией.**

pred = np.argmax(pred, axis=1) + 1

predictions = 0

for i in range(len(pred)):

if pred[i] == y[i][0]:

predictions += 1

print("Accuracy: ", predictions / len(y) \* 100)

Результат выполнения: 97.52%. У логистической регрессией этот показатель был 95.12. Можно сделать вывод, что нейронная сеть справилась лучше с поставленной задачей, т. к. итоговый показатель правильно распознанных образов выше примерно на 2.5%.

**5. Перекодируйте исходные метки классов по схеме one-hot.**

label\_encoder = OneHotEncoder(categories='auto')

y\_one\_hot = label\_encoder.fit\_transform(y.reshape(-1, 1)).toarray()

**6. Реализуйте функцию стоимости для данной нейронной сети.**

**def** softmax\_with\_cross\_entropy(predictions, target\_index):

predictions = predictions - np.max(predictions, axis=1)[:, np.newaxis]

exponents = np.exp(predictions)

softmax\_values = exponents / np.sum(exponents, axis=1)[:, np.newaxis]

loss = -np.log(softmax\_values[range(softmax\_values.shape[0]), target\_index])

loss = np.sum(loss)

dprediction = softmax\_values

dprediction[range(dprediction.shape[0]), target\_index] -= 1

**return** loss, dprediction

**7. Добавьте L2-регуляризацию в функцию стоимости.**

**def** l2\_regularization(W, reg\_strength):

loss = reg\_strength \* np.sum(W\*\*2)

grad = reg\_strength \* 2 \* W

**return** loss, grad

**8. Реализуйте функцию вычисления производной для функции активации.**

def sigmoid\_derivative(z):

return np.multiply(sigmoid(z), 1-sigmoid(z))

**9. Инициализируйте веса небольшими случайными числами.**

def rand\_weights(L\_in, L\_out):

epi = (6 \*\* 1/2) / (L\_in + L\_out) \*\* 1/2

W = np.random.rand(L\_out, L\_in + 1) \* (2 \* epi) - epi

return W

input\_layer\_size = len(X[0]) # 400

num\_labels = len(y\_one\_hot.values[0]) # 10

initial\_Theta1 = rand\_weights(input\_layer\_size, hidden\_layer\_size)

initial\_Theta2 = rand\_weights(hidden\_layer\_size, num\_labels)

initial\_nn\_params = np.append(initial\_Theta1.flatten(), initial\_Theta2.flatten())

**10. Реализуйте алгоритм обратного распространения ошибки для данной конфигурации сети.**

def back\_propagation(nn\_params, hidden\_layer\_size, X, y\_d, lmbda):

input\_layer\_size = len(X[0]) # 400

num\_labels = len(y\_d.values[0]) # 10

initial\_theta1 = np.reshape(nn\_params[:hidden\_layer\_size\*(input\_layer\_size+1)], (hidden\_layer\_size, input\_layer\_size+1), 'F')

initial\_theta2 = np.reshape(nn\_params[hidden\_layer\_size\*(input\_layer\_size+1):], (num\_labels, hidden\_layer\_size+1), 'F')

delta1 = np.zeros(initial\_theta1.shape)

delta2 = np.zeros(initial\_theta2.shape)

m = len(y\_d)

for i in range(X.shape[0]):

ones = np.ones(1)

a1 = np.hstack((ones, X[i]))

z2 = np.dot(a1, initial\_theta1.T)

a2 = np.hstack((ones, sigmoid(z2)))

z3 = np.dot(a2, initial\_theta2.T)

a3 = sigmoid(z3)

d3 = a3 - y\_d.iloc[i, :][np.newaxis, :]

z2 = np.hstack((ones, z2))

d2 = np.multiply(np.dot(initial\_theta2.T, d3.T), sigmoid\_derivative(z2).T[:,np.newaxis])

delta1 = delta1 + np.dot(d2[1:, :], a1[np.newaxis, :])

delta2 = delta2 + np.dot(d3.T, a2[np.newaxis, :])

delta1 /= m

delta2 /= m

delta1[:, 1:] = delta1[:, 1:] + initial\_theta1[:, 1:] \* lmbda / m

delta2[:, 1:] = delta2[:, 1:] + initial\_theta2[:, 1:] \* lmbda / m

return np.hstack((delta1.ravel(order='F'), delta2.ravel(order='F')))

backprop\_params = back\_propagation(initial\_nn\_params, hidden\_layer\_size, X, y\_one\_hot, 0)

**11. Для того, чтобы удостоверится в правильности вычисленных значений градиентов используйте метод проверки градиента с параметром ε = 10-4.**

def gradient\_check(nn\_initial\_params, nn\_backprop\_params, hidden\_layer\_size, X, y\_d, lmbda):

myeps = 0.0001

flattened = nn\_initial\_params

flattenedDs = nn\_backprop\_params

n\_elems = len(flattened)

# Pick ten random elements, compute numerical gradient, compare to respective D's

for i in range(10):

x = int(np.random.rand()\*n\_elems)

epsvec = np.zeros((n\_elems, 1))

epsvec[x] = myeps

cost\_high = cost\_func(flattened + epsvec.flatten(), hidden\_layer\_size, X, y\_d, lmbda)

cost\_low = cost\_func(flattened - epsvec.flatten(), hidden\_layer\_size, X, y\_d, lmbda)

mygrad = (cost\_high - cost\_low) / float(2\*myeps)

print("Element: {0}. Numerical Gradient = {1:.9f}. BackProp Gradient = {2:.9f}."

.format(x, mygrad, flattenedDs[x]))

gradient\_check(initial\_nn\_params, backprop\_params, hidden\_layer\_size, X, y\_one\_hot, 0)

**12. Добавьте L2-регуляризацию в процесс вычисления градиентов.**

model = TwoLayerNet(

W1=np.random.randn(400, 25),

W2=np.random.randn(25, 10),

B1=np.random.randn(25),

B2=np.random.randn(10),

reg=1e-4

)

**13. Проверьте полученные значения градиента.**

backprop\_params = back\_propagation(initial\_nn\_params, hidden\_layer\_size, X, y\_one\_hot, lmbda)

gradient\_check(initial\_nn\_params, backprop\_params, hidden\_layer\_size, X, y\_one\_hot, lmbda)

Результат выполнения:

Element: 986. Numerical Gradient = -0.000000734. BackProp Gradient = -0.000000734.

Element: 1521. Numerical Gradient = 0.000000711. BackProp Gradient = 0.000000711.

Element: 2965. Numerical Gradient = 0.000085671. BackProp Gradient = 0.000085671.

Element: 3249. Numerical Gradient = -0.000711440. BackProp Gradient = -0.000711440.

Element: 9632. Numerical Gradient = 0.000002351. BackProp Gradient = 0.000002351.

Element: 8061. Numerical Gradient = -0.000000715. BackProp Gradient = -0.000000715.

Element: 3879. Numerical Gradient = -0.003488895. BackProp Gradient = -0.003488895.

Element: 7718. Numerical Gradient = 0.000508230. BackProp Gradient = 0.000508230.

Element: 4952. Numerical Gradient = -0.001355993. BackProp Gradient = -0.001355993.

Element: 6969. Numerical Gradient = 0.000061777. BackProp Gradient = 0.000061777.

**14. Обучите нейронную сеть с использованием градиентного спуска или других более эффективных методов оптимизации.**

theta\_opt = opt.fmin\_cg(maxiter=30, f=cost\_func, x0=initial\_nn\_params, fprime=back\_propagation, args=(hidden\_layer\_size, X, y\_one\_hot, lmbda))

print(theta\_opt)

theta1\_opt = np.reshape(theta\_opt[:hidden\_layer\_size\*(input\_layer\_size+1)], (hidden\_layer\_size, input\_layer\_size+1), 'F')

theta2\_opt = np.reshape(theta\_opt[hidden\_layer\_size\*(input\_layer\_size+1):], (num\_labels, hidden\_layer\_size+1), 'F')

**15. Вычислите процент правильных классификаций на обучающей выборке.**

pred = h0x(X, [theta1\_opt, theta2\_opt])

pred = np.argmax(pred, axis=1) + 1

predictions = 0

for i in range(len(pred)):

if pred[i] == y[i][0]:

predictions += 1

print("Accuracy: ", predictions / len(y) \* 100)

Результат выполнения:

'Accuracy: ', 93.84

**16. Визуализируйте скрытый слой обученной сети.**

W1\_trained = model.fc1.W.value

W1\_trained\_reshaped = W1\_trained.reshape((20, 20, 25)).transpose()

**for** W **in** W1\_trained\_reshaped[:5]:

plt.imshow(W.transpose())

plt.axis('off')

plt.colorbar()

plt.show()

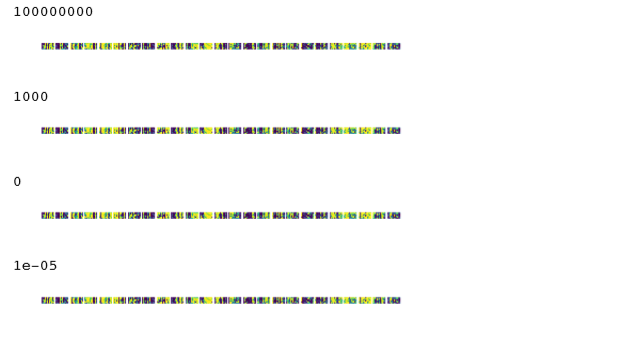


Рисунок 1 – визуализация скрытого слоя при lambda=1

**17. Подберите параметр регуляризации. Как меняются изображения на скрытом слое в зависимости от данного параметра?**

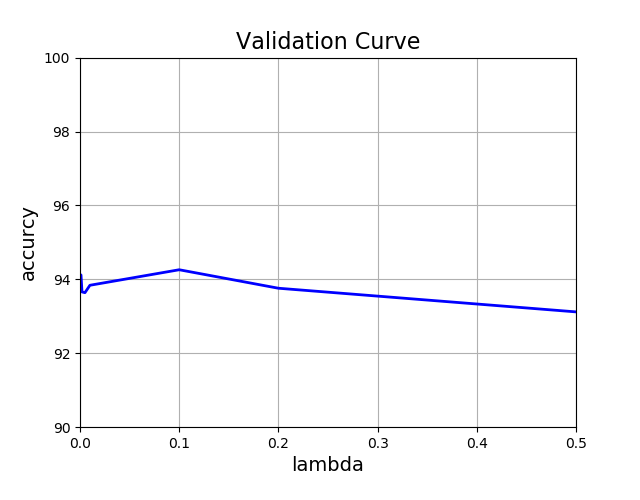


Рисунок 2 – точность предсказаний при различных lambda (на рисунке видно, что наибольшая точность достигается при lambda=0.1)

Мы видим, что по краям изображений становится мало информации, а значит веса там близки к нулю.

**Вывод**

Нейронные сети используются для решения сложных задач, которые требуют аналитических вычислений подобных тем, что делает человеческий мозг. Самыми распространенными применениями нейронных сетей является: Классификация Предсказание Распознавание Нейрон — это вычислительная единица, которая получает информацию, производит над ней простые вычисления и передает ее дальше. Они делятся на три основных типа: входной, скрытый и выходной. Также есть нейрон смещения и контекстный нейрон о которых мы поговорим в следующей статье. В том случае, когда нейросеть состоит из большого количества нейронов, вводят термин слоя. Соответственно, есть входной слой, который получает информацию, n скрытых слоев, которые ее обрабатывают и выходной слой, который выводит результат. У каждого из нейронов есть 2 основных параметра: входные данные (input data) и выходные данные (output data). В случае входного нейрона: input=output. В остальных, в поле input попадает суммарная информация всех нейронов с предыдущего слоя, после чего, она нормализуется, с помощью функции активации (пока что просто представим ее f(x)) и попадает в поле output.