Notes regarding lab format

We will use Matlab Livescript for this lab. Livescript allows switching between text and code cells.

You will find the entire lab manual in this file. Some exercises require you to write a text answer, others require you to write code. You should not define functions inside this file. Instead save functions to the functions folder and call them from the code cells in this notebook.

Your finished lab report should be a .zip-file containing the data folder, your functions folder, any output images you might have saved and this livescript file.

Since we need to access the functions and data folder the first step is to add these two locations to MATLAB's path.

```
addpath('./functions');
addpath('./data');
```

Lab 2: Learning and convolutional neural networks

2.1 Learning a Linear Classifier

In this part, we will try to learn a linear classifier for blood cell detection. Note that the classifier could also be viewed as a minimal neural network consisting of three parts: a scalar product node (or fully-connected node), a constant (or bias) term and a logistic sigmoid function. To find good parameters we will try to minimize the negative log-likelihood over a small training set.

The output from our classifier is a probability p for the input patch being centered at a cell centre. The sigmoid function will make sure that $0 \leq p \leq 1$. To be more precise the output is

$$p = \frac{e^y}{1 + e^y} \quad \text{where} \quad y = \mathrm{I} \cdot \omega + \omega_0$$

Instead of testing a bunch of manually chosen w's and w_0's, we will try to learn good values for all the parameters. This requires training examples, that you find in cell_data.mat.

Ex 2.1

Load the data using

```
load cell_data.mat
```

It loads a structure, cell_data, with two fields, `fg_patches` and `bg_patches`, corresponding to positive (centered blood cells) negative examples respectively.

Ex 2.2

Create two new variables, examples and labels. The variable examples should be a cell structure containing all the patches (both positives and negatives) and labels should be an array with the same number of elements such that `labels(i) = 1` if `examples{i}` is a positive example, and `labels(i) = 0` otherwise.

```
% Your code here
examples = [cell_data.fg_patches, cell_data.bg_patches];
labels = [zeros(1, size(cell_data.fg_patches,2)), ones(1,size(cell_data.bg_patches,2))];
```

Ex 2.3

Split the data into training, `(examples_train, labels_train)`, and validation, `(examples_val, labels_val)`. The two should have a similar structure to examples and labels. Write on the report which percentage of the data you used for validation. Also, the splitting of the data into the two sets should be done in a random manner, for example, using randperm.

Percentage of data used for training: 80%

Percentage of data used for validation: 20%

```
% Your code here
train_perc = 0.8;
N = size(examples,2);

train_size = round(N*train_perc);

indexes = randperm(N);

examples_train = examples(indexes(1:train_size));
labels_train = labels(indexes(1:train_size));

examples_val = examples(indexes(train_size+1:end));
labels_val = labels(indexes(train_size+1:end));
```

2.2 Training the classifier

We will try to find parameters that minimize the negative log-likelihood on the training data. More precisely,

$$L(\theta) = \sum_{i \in S_+} -\ln(p_i) + \sum_{i \in S_-} -\ln(1 - p_i) = \sum_i L_i(\theta)$$

where p_i refers to the classifier output for the ith training example. As in the lectures we will refer to the terms here as the partial loss L_i.

Before doing the next exercise, you need to work out how to compute the gradient of the partial loss L_i.

Ex 2.4

Make a function

```
[wgrad, w0grad] = partial_gradient(w, w0, example_train, label_train)
```

that computes the derivatives of the partial loss L_i with respect to each of the classifier parameters. Let the output wgrad be an array of the same size as the weight image, w (and let w0grad be a number).

At each iteration of stochastic gradient descent, a training example, i, is chosen at random. For this example the gradient of the partial loss, L_i , is computed and the parameters are updated according to this gradient. The most common way to introduce the randomness is to make a random reordering of the data and then going through it in the new order. One pass through the data is called an epoch.

Ex 2.5

Make a function

```
[w, w0] = process_epoch(w, w0, lrate, examples_train, labels_train)
```

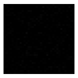
that performs one epoch of stochastic gradient descent

Ex 2.6

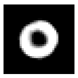
Initialize `w = s * randn(35,35);`, with `s = 0.01` and `w0 = 0;` and run 5 epochs on your training examples. Plot w after each epoch (or after each iteration if you are curious), to get a sense of what is happening. Also, try using different `s={10,1,0.1,0.01}` and plot w after 5 epochs for each value of s. Include on the report visualizations of w for the different values of s, along with an written explanation of what is happening.

```
% your code here
s = 0.1;
w = s * randn(35,35);
w0 = 0;
n_epochs = 5;
lrate = 0.1;
for s = [0.01, 10, 1, 0.1]
    w = s * randn(35, 35);
    w0 = 0;
    disp(['s =  ' num2str(s)])
    disp(['Initial w: ']), figure, colormap gray, imshow(w)
    for n = 1:n_epochs
        [w, w0] = process_epoch(w, w0, lrate, examples_train, labels_train);
    end
    disp(['Final w:']), figure, colormap gray, imshow(w)
end
```

```
s =  0.01
Initial w:
```



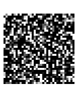
Final w:



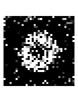
s = 10
Initial w:

Final w:

s = 1
Initial w:

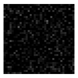


Final w:

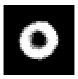


s = 0.1

```
Initial w:
```



```
Final w:
```



Your written answer here:

s determines the magnitude of the random values used to initialize the w matrix. As we can see, when s = 0.01 is pretty good for our case. If the magnitude of the initial values in w is too big, the optimization process can become unstable and overshoot the optimal solution.

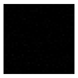
Exercise 2.7

As said before, at each iteration of stochastic gradient descent, a training example is chosen at random. Check what happens to w after 5 epochs when that training example is not chosen randomly but in sequence, i.e. first `i = 1`, then `i = 2`, and so on. Include on the report a visualization of w for this case. (Don't forget to change back your function to a random choice of i after this exercise).

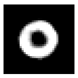
You can display the resulting filters inside this livescript notebook using the function imagesc.

```
% Your code here
%for2.7, change the process_epoch code to linspace
lrate = 0.1;
epochs = 5;
for s = [0.01, 10, 1, 0.1]
    w = s * randn(35, 35);
    w0 = 0;
    disp(['s =  ' num2str(s)])
    disp(['Initial w: ']), figure, colormap gray, imshow(w)
    for n = 1:epochs
        [w, w0] = process_epoch(w, w0, lrate, examples_train, labels_train);
    end
    disp(['Final w:']), figure, colormap gray, imshow(w)
end
```

```
s =  0.01
Initial w:
```



```
Final w:
```

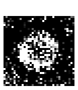


```
s =  10
Initial w:
```

```
Final w:
```

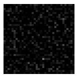
```
s =  1
Initial w:
```

```
Final w:
```

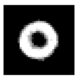


```
s =  0.1
```

```
Initial w:
```



```
Final w:
```



Ex 2.8

Make a function

`predicted_labels = classify(examples_val,w,w0);`

that applies the classifier to the example data. After that, use it on `examples_train` and examples_val and check how much accuracy it gets for each by comparing the predicted labels with `labels_train` and `labels_val` respectively. Write on your report the highest accuracy you were able to achieve in the training and validation data. Hint: train the classifier for longer than 5 epochs to make sure that it converges.

When you have defined the function run the following code.

```
s = 0.001;
w = s * randn(35,35);
w0 = 0;
n_epochs = 500;
lrate = 0.01;

for i=1:n_epochs

    [w,w0] = process_epoch(w,w0,lrate,examples_train,labels_train);

end
predicted_labels = classify(examples_val,w,w0);
correct = sum(predicted_labels==labels_val);
accuracy = 100*correct/length(labels_val);
fprintf(["Validation Accuracy: "]+accuracy+["%%"])
```

```
Validation Accuracy: 91.25%
```

Write the highest accuracy you were able to get here:

91.25%

Ex 2.9

The data for training this classifier consists on only 400 examples (less if you consider that you have split it into training and validation). To achieve higher accuracy it might be useful to perform some data augmentation before the training. In this exercise you will increase the number of elements in the training examples by M times. Make a function

```
[examples_train_aug,labels_train_aug] = augment_data(examples_train,labels_train,M)
```

that takes each sample of the original training data and applies M random rotations (you can use Matlab function imrotate), from which result M new examples. Store these new examples in examples_train_aug and their corresponding labels in labels_train_aug. Train the classifier with this augmented data and write on your report the new values for accuracy on the training and validation examples.

```
% Your code here
s = 0.001;
w = s * randn(35,35);
w0 = 0;
n_epochs = 500;
lrate = 0.01;
M = 3;
[examples_train_aug,labels_train_aug] = augment_data(examples_train,labels_train,M);

for i=1:n_epochs

    [w,w0] = process_epoch(w,w0,lrate,examples_train_aug,labels_train_aug);

end

predicted_labels = classify(examples_val,w,w0);
correct = sum(predicted_labels==labels_val);
accuracy = 100*correct/length(labels_val);
fprintf(("Validation Accuracy: ")+accuracy+["%%"])
```

```
Validation Accuracy: 92.5%
```

Write the highest accuracy you were able to get here:

95%

2.3 Convolutional neural networks

In the last part, your task is to train convolutional neural networks using Matlab.

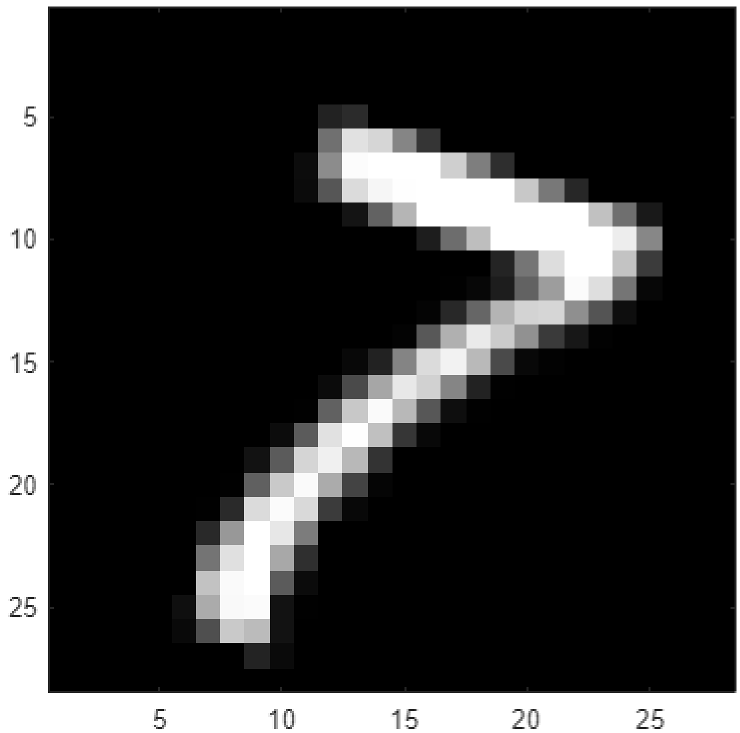
Ex 2.10

Run the following cell

```
[imgs, labels] = digitTrain4DArrayData;
```

to load a dataset of images of digits into Matlab. (Make sure you have the Deep Learning toolbox installed). You will find the 5000 digit images in imgs. Plot a few of them to see what the data looks like.

```
N = 3210; % We have 5000 images so any number between 1 and 5000 works.
img = imgs(:,:,:,N);
imagesc(img), axis image, colormap gray
```



The next step is to define a network for classification. In Matlab, you do this by simply giving an array of the layers. For example, this would be a linear classifier similar to the one you trained for cells:

layers = [

imageInputLayer([35 35 1]);

fullyConnectedLayer(1);

softmaxLayer();

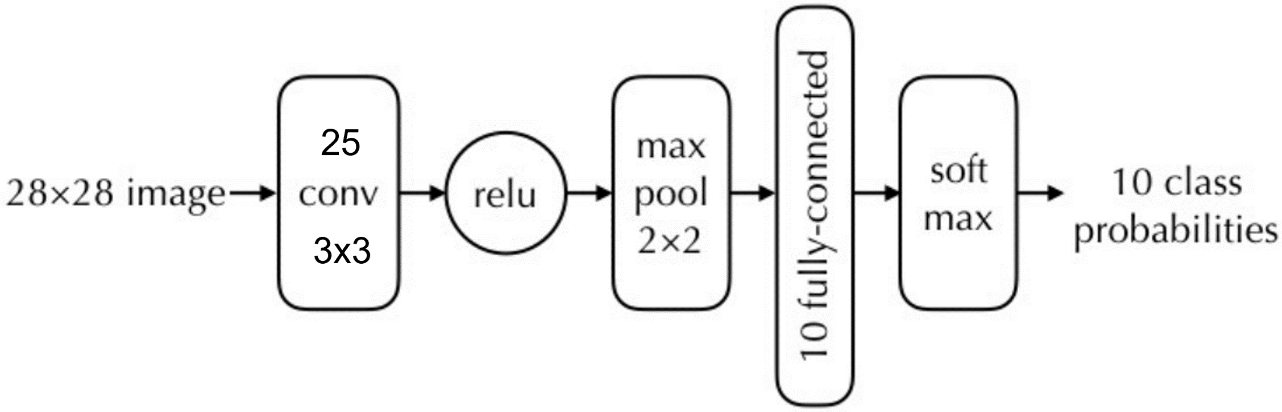
classificationLayer()];

Ex 2.11

Make a function

`layers = basic_cnn_classifier()`

that implements the following network in Matlab:



Apart from the layers above the functions `convolution2dLayer`, `reluLayer` and `maxPooling2dLayer` will be useful. Note that you have to set the *stride* for max pooling to 2 to get the expected downsampling.

When you have written the function run the following cell:

```
layers = basic_cnn_classifier()
```

```
layers = 
  7×1 Layer array with layers:

     1   ''   Image Input             28×28×1 images with 'zerocenter' normalization
     2   ''   Convolution             25 3×3 convolutions with stride [1  1] and padding [0  0  0  0]
     3   ''   ReLU                    ReLU
     4   ''   Max Pooling             2×2 max pooling with stride [2  2] and padding [0  0  0  0]
     5   ''   Fully Connected         10 fully connected layer
     6   ''   Softmax                 softmax
     7   ''   Classification Output   crossentropyex
```

Ex 2.12

Create a set of training options telling Matlab to use stochastic gradient descent with momentum (SGDM), for the optimization:

`options = trainingOptions('sgdm');`

Now train the network (using default parameters) by running

net = trainNetwork(imgs, labels, layers, options)

```
options = trainingOptions('sgdm');
net = trainNetwork(imgs, labels, layers, options);
```

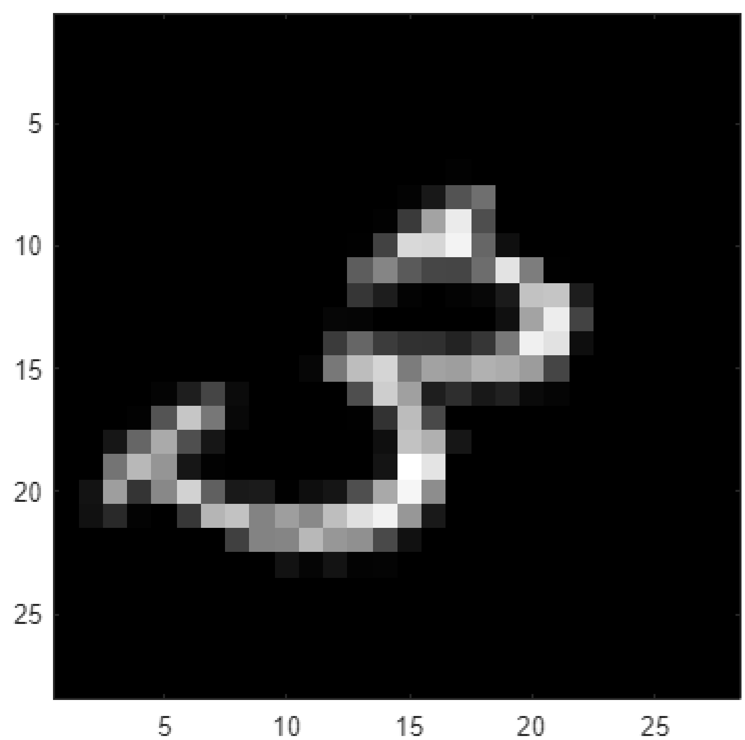
```
Training on single CPU.
Initializing input data normalization.
|========================================================================================|
|  Epoch  |  Iteration  |  Time Elapsed  |  Mini-batch  |  Mini-batch  |  Base Learning  |
|         |             |   (hh:mm:ss)   |   Accuracy   |     Loss     |      Rate       |
|========================================================================================|
|       1 |           1 |       00:00:00 |        7.81% |       2.3006 |          0.0100 |
|       2 |          50 |       00:00:01 |       46.88% |       1.8064 |          0.0100 |
|       3 |         100 |       00:00:03 |       64.06% |       1.2163 |          0.0100 |
|       4 |         150 |       00:00:04 |       60.16% |       1.0487 |          0.0100 |
```

```
|       6 |        200 |      00:00:05 |      66.41% |       1.0300 |          0.0100 |
|       7 |        250 |      00:00:06 |      67.19% |       1.0983 |          0.0100 |
|       8 |        300 |      00:00:08 |      63.28% |       0.9246 |          0.0100 |
|       9 |        350 |      00:00:09 |      74.22% |       0.9170 |          0.0100 |
|      11 |        400 |      00:00:11 |      70.31% |       0.8786 |          0.0100 |
|      12 |        450 |      00:00:12 |      81.25% |       0.6564 |          0.0100 |
|      13 |        500 |      00:00:14 |      77.34% |       0.7490 |          0.0100 |
|      15 |        550 |      00:00:15 |      84.38% |       0.6070 |          0.0100 |
|      16 |        600 |      00:00:16 |      83.59% |       0.5303 |          0.0100 |
|      17 |        650 |      00:00:18 |      84.38% |       0.4480 |          0.0100 |
|      18 |        700 |      00:00:19 |      91.41% |       0.4042 |          0.0100 |
|      20 |        750 |      00:00:20 |      91.41% |       0.4476 |          0.0100 |
|      21 |        800 |      00:00:21 |      91.41% |       0.4031 |          0.0100 |
|      22 |        850 |      00:00:22 |      96.09% |       0.2977 |          0.0100 |
|      24 |        900 |      00:00:24 |      94.53% |       0.2597 |          0.0100 |
|      25 |        950 |      00:00:25 |      96.88% |       0.1874 |          0.0100 |
|      26 |       1000 |      00:00:26 |      96.09% |       0.2050 |          0.0100 |
|      27 |       1050 |      00:00:28 |      98.44% |       0.1973 |          0.0100 |
|      29 |       1100 |      00:00:29 |      95.31% |       0.2366 |          0.0100 |
|      30 |       1150 |      00:00:31 |      98.44% |       0.1460 |          0.0100 |
|      30 |       1170 |      00:00:31 |      97.66% |       0.1474 |          0.0100 |
|=====================================================================================|
Training finished: Max epochs completed.
```

Ex 2.13

Try the network on a few of the training images. You can use `net.predict(img)` to get the ten output probabilities or `net.classify(img)` to get the most probable class.

```
img = imgs(:,:,:,11);
imagesc(img), axis image, colormap gray
```



```
net.predict(img)
```

```
ans = 1×10 single row vector
    0.0001    0.0001    0.0612    0.8503    0.0004    0.0670    0.0088    0.0023 ···
```

```
net.classify(img)
```

```
ans = categorical
     3
```

Ex 2.14

Work out how many trainable parameters your network contains. Include the answer in your submission. If you explore what the data structure net actually contains, you can find the answer there as well. Note that the convolution layer does not use padding so the output from the convolution layer is smaller than the input.

Write your answer here:

convolution2dLayer: each of the 25 filters has 3x3x1=9 parameters. So, the total number of parameters for this layer is 25 x (9+1) = 250.

fullyConnectedLayer: each of the 10 neurons has 26x26x25/4=4225 inputs (a stride of 2 in the max pooling layer) . So, the total number of parameters for this layer is 10 x 4225+10 = 42260.

Total: 42510

Ex 2.15

Matlab prints a lot of output, for example the accuracy on the training set. Recall from the lectures that this number is not very good for judging the quality of a classifier. Instead we should save a subset of the data as a validation set, that we can use to evaluate the trained network.

Divide the data into training `imgs` and `labels` into new variables `imgs_train, labels_train, imgs_val, labels_val`.

```
split_perc = 0.8;
N = length(imgs);
train_size = round(split_perc*N);
test_size = N - train_size;
imgs_train = imgs(:,:,:,1:train_size);
labels_train = labels(1:train_size);
imgs_val = imgs(:,:,:,train_size+1:end);
labels_val = labels(train_size+1:end);
```

Make a function

`net = train_classifier(layers, imgs_train, labels_train, imgs_val, labels_val)`

that runs a few epochs of training and then evaluates the accuracy on the validation set. In this case, Matlab has given us a separate test set, so we don't have to save images for that purpose. *Matlab hints:* You can run multiple images at once by stacking them along the fourth dimension. If you want to continue training the same network you can run

```
net = train_classifier(net.Layers, imgs_train, labels_train, imgs_val, labels_val)
```

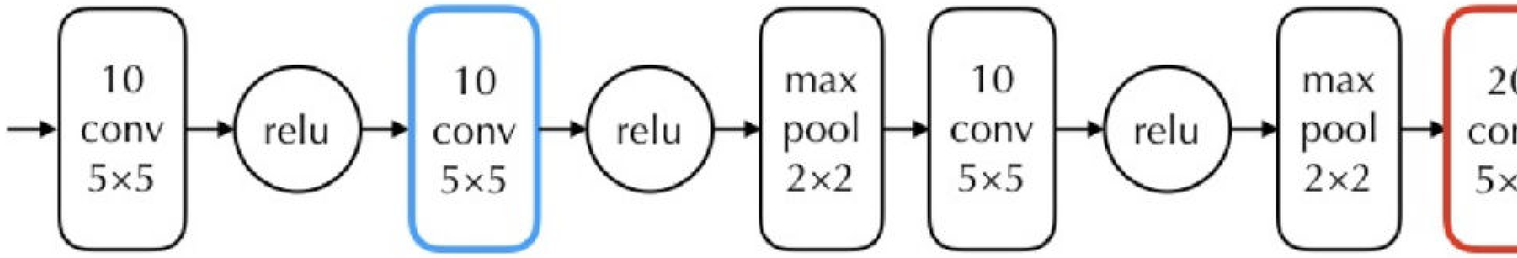
```
net = train_classifier(net.Layers, imgs_train, labels_train, imgs_val, labels_val)
```

```
Training...
The network achieved an accuracy of: 98%
net =
  SeriesNetwork with properties:

         Layers: [7×1 nnet.cnn.layer.Layer]
     InputNames: {'imageinput'}
    OutputNames: {'classoutput'}
```

Ex 2.16

To run a convolutional neural network we have to perform a massive amount of computations. Hence it is very important to consider the computational load when designing a network. For the network below, compare (roughly) the time consumption of the blue and the red layers. You can ignore effects of padding. Include your answer and your motivation in your submission.



Write your answer here:

For trainable parameters:

Blue convolutional layerh has (10*5*5+1)*10=2510 parametrs.

Red convolutional layer has (10*5*5+1)*20=5020.

For input parameters, it depends on the size of input image. Between red and blue layer, there exists two max pooling layer, which would decrease the size of input parameters with 16X.

For the time consumption, assume the input image for the blue layer has the size of N*N and for the red layer it comes to (N/4)*(N/4),we can compute roughly as follows:

The blue layer: the time consumption=c1*5*5*10*N*N=6250N^2

The red layer: the time consumption=c2*5*5*20*(N*N/16)=125N^2

It shows that the red layer is faster than the blue layer.

Ex 2.17

Replace the blue box of the network in the figure above by a sequence of two layers of 10 3 × 3 convolutional filters. What changes in terms of network parameters, time consumption and accuracy? Again, you are not supposed to implement the network.

Write your answer here:

Parameters: (10*3*3+1)*10*2 = 1820, which is less than one 10 5*5 layer. The reduction of parameters might potentially decrease the time consumption for both forward and backward propagation, as fewer parameters means a faster training process.

For time consumption,the blue layer : the red layer=5*5 : 2*3*3=25 : 18. So the red layer is faster than the blue layer

For accuracy, the model seems to get deeper and have more patterns which might improve the overall accuracy of the model. It depends on the experiment.

Ex. 2.18

Make a copy of basic_cnn_classifier.m and name it better_cnn_classifier.m. Try modifying the network by adding more layers. Also experiment with the training options. How much can you improve the results?

```
layers = better_cnn_classifier()
```

```
layers = 
  12×1 Layer array with layers:

     1   ''   Image Input             28×28×1 images with 'zerocenter' normalization
     2   ''   Convolution             20 5×5 convolutions with stride [1  1] and padding [0  0  0  0]
     3   ''   ReLU                    ReLU
     4   ''   Max Pooling             2×2 max pooling with stride [2  2] and padding [0  0  0  0]
     5   ''   Convolution             30 3×3 convolutions with stride [1  1] and padding [0  0  0  0]
     6   ''   ReLU                    ReLU
     7   ''   Max Pooling             2×2 max pooling with stride [2  2] and padding [0  0  0  0]
     8   ''   Fully Connected         100 fully connected layer
     9   ''   ReLU                    ReLU
    10   ''   Fully Connected         10 fully connected layer
    11   ''   Softmax                 softmax
    12   ''   Classification Output   crossentropyex
```

```
net = train_classifier(layers, imgs_train, labels_train, imgs_val, labels_val)
```

```
The network achieved an accuracy of: 99.8%
net = 
  SeriesNetwork with properties:

         Layers: [12×1 nnet.cnn.layer.Layer]
     InputNames: {'imageinput'}
    OutputNames: {'classoutput'}
```

Ex 2.19

Load the builtin test images using

```
[imgs_test, labels_test] = digitTest4DArrayData;
```

Run the network on all the test images. You apply the network to an image using

pred = net.classify(image)

Compute precision and recall for each of the 10 classes and include these in your submission. The definitions of precision and recall can be found in the lecture notes chapter 4.

```
[imgs_test, labels_test] = digitTest4DArrayData;
pred = net.classify(imgs_test);
for class = 0:9
    labels_class_idx = labels_test == num2str(class);
    correct_predictions = nnz(pred(labels_class_idx) == num2str(class));
    precision = correct_predictions / nnz(pred == num2str(class));
    recall = correct_predictions / length(labels_test(labels_class_idx));
    disp(['Precision ' num2str(class) ': ' num2str(precision)]);
    disp(['Recall  ' num2str(class) ':  ' num2str(recall)]);
end
```

```
Precision 0: 1
Recall  0:  0.99
Precision 1: 0.99592
Recall  1:  0.976
Precision 2: 0.98617
Recall  2:  0.998
Precision 3: 0.97642
Recall  3:  0.994
Precision 4: 1
Recall  4:  0.998
Precision 5: 0.99404
Recall  5:  1
Precision 6: 0.99401
Recall  6:  0.996
Precision 7: 0.992
Recall  7:  0.992
Precision 8: 0.99598
Recall  8:  0.992
Precision 9: 0.99599
Recall  9:  0.994
```

Ex 2.20

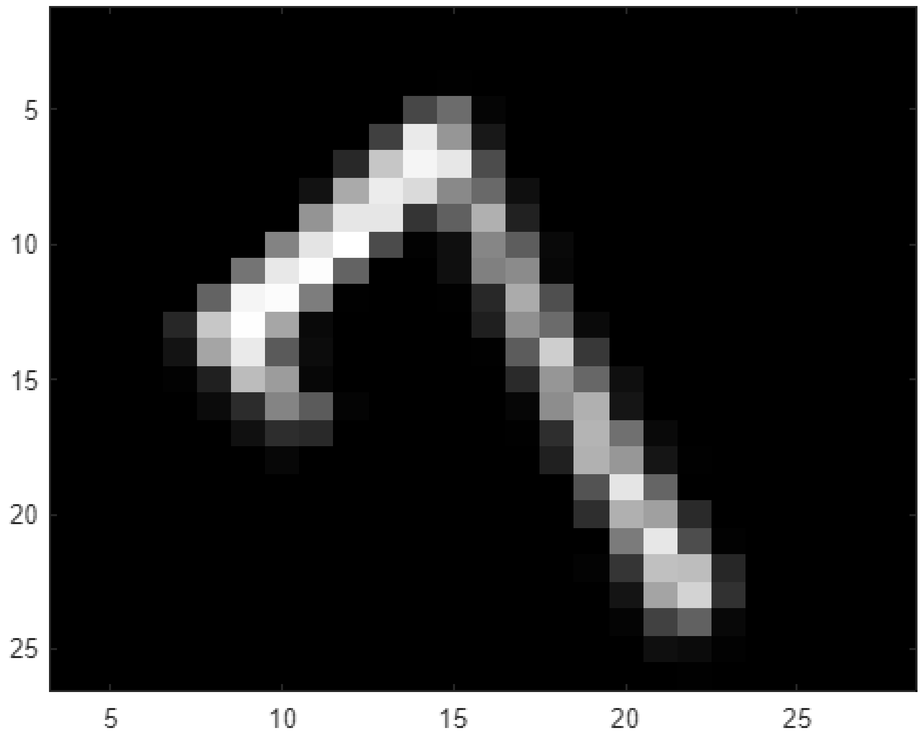
Save three of the failure cases with names indicating what digit they were mistaken for. Include these in your submission. You can use imwrite(img, 'mistaken_as_5.png') to save an image if it is correctly scaled. Have a look at the file before submitting it, so it looks right.

You can load and display the images here using imagesc:

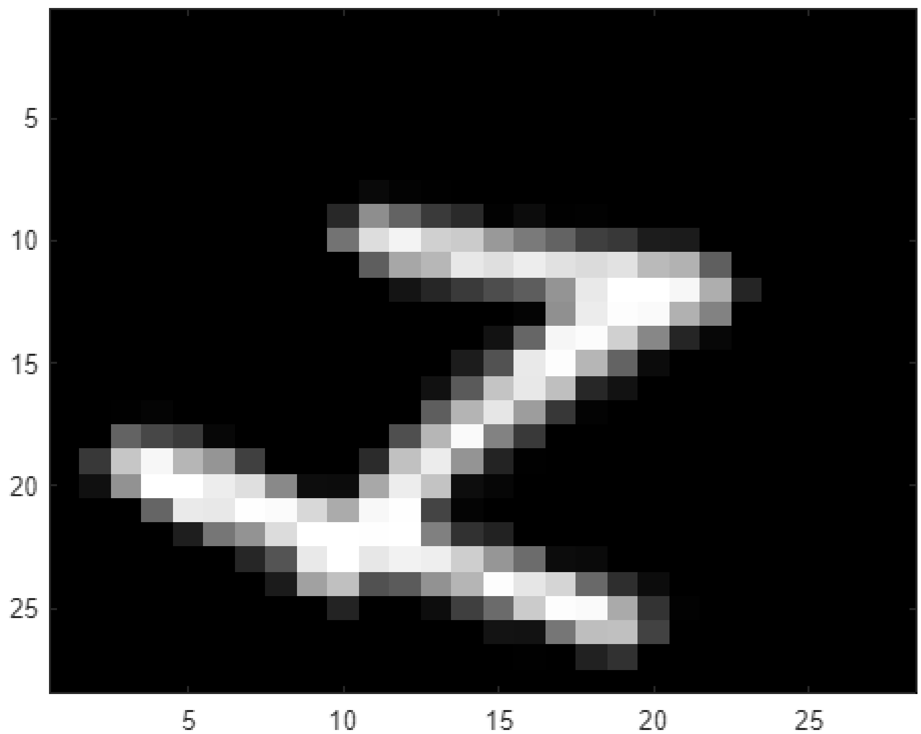
```
n_fails = 3;
fail = find(pred ~= labels_test);
idx = randperm(length(fail),n_fails);
fail_idx = fail(idx);
for i=1:3
    label = labels_test(fail_idx(i));
    predicted_label = pred(fail_idx(i));
```

```
    img = imgs_test(:,:,:,fail_idx(i));
    disp(['Real label: ' label])
    disp(['Predicted label: ' predicted_label])
    figure,imagesc(img),colormap gray
    img_name = strcat('./data/',string(label), '_mistaken_as_', string(predicted_label),'.png'
    imwrite(img, img_name);
end
```

```
Real label:      7
Predicted label:      5
```




```
Real label:      1
Predicted label:      2
```




```
Real label:      0
Predicted label:      8
```