Chapter07 Back Propagation

1. To compute the gradients efficiently, we use back propagation
2. $\frac{\partial C}{\partial W} = \frac{\partial C}{\partial Z} \cdot \frac{\partial Z}{\partial W}$
   a) Forward Pass: Compute $\frac{\partial Z}{\partial W}$ for all parameters

      The result is the input connected by the weight

   b) Backward Pass: Compute $\frac{\partial C}{\partial Z}$ for all activation function inputs z

      $\frac{\partial C}{\partial Z} = \frac{\partial C}{\partial a} \cdot \frac{\partial a}{\partial Z} = \frac{\partial C}{\partial a} \cdot \sigma'(z)$

Chapter08 Keras

1. Keras is the interface of Tensorflow or Theano
2. Code:
   a) model = Sequential()
   b) model.add(Dense(input_dim=28*28,units = 500,activation = 'relu'))
   c) model.add(Dense(units = 500, activation = 'relu'))
   d) model.add(Dense(units = 10, activation = 'softmax'))
   e) model.compile(loss = 'categorical_crossentropy', optimizer = 'adam', metrics = ['accuracy'])
   f) model.fit(x_train, y_train, batch_size = 100, epochs = 20)
3. Batch size influences both speed and performance, usually with the increasing of batchsize, speed will be faster, and performance will drop.
4. Mini-batch is faster than stochastic gradient descent, the reason is that the use of GPU can speed up the matrix operation