COMP9444
Neural Networks and Deep Learning
Term 2, 2025

Week 5 Tutorial Solutions

This page was last updated: 07/06/2025 21:59:02

1. **Convolutional Network Architecture**

   One of the early papers on Deep Q-Learning for Atari games (Mnih et al, 2013) contains this description of its Convolutional Neural Network:

   > "The input to the neural network consists of an 84 × 84 × 4 image. The first hidden layer convolves 16 8 × 8 filters with stride 4 with the input image and applies a rectifier nonlinearity. The second hidden layer convolves 32 4 × 4 filters with stride 2, again followed by a rectifier nonlinearity. The final hidden layer is fully-connected and consists of 256 rectifier units. The output layer is a fully-connected linear layer with a single output for each valid action. The number of valid actions varied between 4 and 18 on the games we considered."

   For each layer in this network, compute the number of

   a. weights per filter in this layer (including bias)
   b. width and height of layer (convolutional layers)
   c. neurons in this layer
   d. connections into the neurons in this layer
   e. independent parameters in this layer

   Four consecutive images from the game are presented to the network simultaneusly. You should assume the input images are gray-scale, there is no padding, and there are 18 valid actions (outputs).

First Convolutional Layer:

$J = K = 84,\ L = 4,\ M = N = 8,\ P = 0,\ s = 4$

weights per filter: $1 + M \times N \times L = 1 + 8 \times 8 \times 4 = 257$
width and height of layer: $1 + (J - M) / s = 1 + (84 - 8) / 4 = 20$
neurons in layer: $20 \times 20 \times 16 = 6400$
connections: $20 \times 20 \times 16 \times 257 = 1644800$
independent parameters: $16 \times 257 = 4112$

Second Convolutional Layer:

$J = K = 20,\ L = 16,\ M = N = 4,\ P = 0,\ s = 2$

weights per filter:	$1 + M \times N \times L = 1 + 4 \times 4 \times 16 = 257$
width and height of layer:	$1 + (J - M) / s = 1 + (20 - 4) / 2 = 9$
neurons in layer:	$9 \times 9 \times 32 = 2592$
connections:	$9 \times 9 \times 32 \times 257 = 666144$
independent parameters:	$32 \times 257 = 8224$

Fully Connected Layer:

weights per neuron:	$1 + 2592 = 2593$
neurons in layer:	256
connections:	$256 \times 2593 = 663808$
independent parameters:	663808

Output Layer:

weights per neuron:	$1 + 256 = 257$
neurons in layer:	18
connections:	$18 \times 257 = 4626$
independent parameters:	4626

---

2. **Weight Initialization**

Briefly describe the problem of vanishing or exploding gradients, and how Weight Initialization can be used to prevent it.

The differentials in a deep neural network tend to grow according to this equation:

$$\mathrm{Var}[\partial/\partial x] \simeq (\Pi_{1 \le i \le \mathrm{D}}\ \mathrm{G}_1 n_{(i+1)}\ \mathrm{Var}[w^{(i)}])\ \mathrm{Var}[\partial/\partial z]$$

where $w^{(i)}$ are the weights at layer i, $n_{(i+1)}$ is the number of weights fanning out from each node in layer i, and G_1 estimates the average value of the derivative of the transfer function. If the weights are initialized so that the factor in parentheses corresponding to each layer is approximately 1, then the differentials will remain in a healthy range. Otherwise, they may either grow or vanish exponentially.

---

3. **Batch Normalization**

Briefly describe the Batch Normalization algoritm.

The mean and variance of the activations $x_k^{(i)}$ at layer i over a batch of training items are estimated or pre-computed, and normalized activations are calculated for each node:

$$\hat{x}_k^{(i)} = (x_k^{(i)} - \mathrm{Mean}[x_k^{(i)}])/(\sqrt{\mathrm{Var}[x_k^{(i)}]})$$

These activations are then shifted and rescaled by $y_k^{(i)} = \beta_k^{(i)} + \gamma_k^{(i)}\,\hat{x}_k^{(i)}$, where $\beta_k^{(i)}, \gamma_k^{(i)}$ are additional parameters to be learned by backpropagation.

---

4. **Residual and Dense Networks**

   Explain the difference between a Residual Network and a Dense Network.

A Residual Network includes "skip" connections which bypass each pair of consecutive layers. These intermediate layers therefore compute a residual component, which is added to the output from previous layers and corrects their errors, or provides additional details which they were not powerful enough to compute.

A Dense Network is built from densely connected blocks, separated by convolution and pooling layers. Within a dense block, each layer is connected by shortcut connections to all preceding layers.

---

5. **Introduction, Literature Review, and Dataset Analysis**

   This Week's checkpoint task is to read relevant paper(s) related to the project and perform basic dataset analysis. We believe that irrespective of what method(s) or model(s) you will be applying, or what role you have been assigned within the group, it is important for every team member to have a good understanding of the project, related work, dataset and analysis.

   a. Download the **Presentation Template** available at WebCMS → Project → Project Presentation Template

   b. Draft **Motivation** and **Problem Statement** on the slide deck.
   Remember – it's just a draft so don't worry too much about formatting.
   You can keep revising until Week 10 when you have to give the project presentation as a team.

   c. Read at least one relevant paper as part of the literature review and provide a summary on the slide deck under **Literature Review**.
   To get the most out of this activity, you can co-ordinate with your fellow team members so that each member reads different papers in detail.
   In this way, your team will have a summary of four or five papers as an outcome of this activity.

d. Do some basic **Data Analysis**; in particular, try to identify in what format data is given, how many samples, how many classes, what is the distribution of samples for different classes (or categories), dataset quality in terms of diversity, physical characteristics, etc. Briefly provide these details on the slide deck.

e. **Show your draft slide deck to your tutor** and get feedback.

6. **Other Questions**

   Any questions or discussion about image processing, the group project, societal impact, or any other aspect of the course.
