

Assignment Project Exam Help

Announcements

Add WeChat powcoder

Reminder: ps3 due tonight 10/8 at midnight (Boston)

- ps4 out today, due 10/15 (1 week)
- ps3 self-grading form out Monday, due 10/19
<https://powcoder.com>
- Grades for ps1 & ps2 are being posted to blackboard (by Monday)
- Midterm 10/22 – have to finish test once began, should have blank paper that you will submit work/steps for a solution



Assignment Project Exam Help

Add WeChat powcoder

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Neural Networks IV

Recurrent Networks

Assignment Project Exam Help

Today: Outline

Add WeChat powcoder

- **Recurrent networks:** forward pass, backward pass
- **NN training strategies:** loss functions, dropout, etc.

<https://powcoder.com>

Add WeChat powcoder

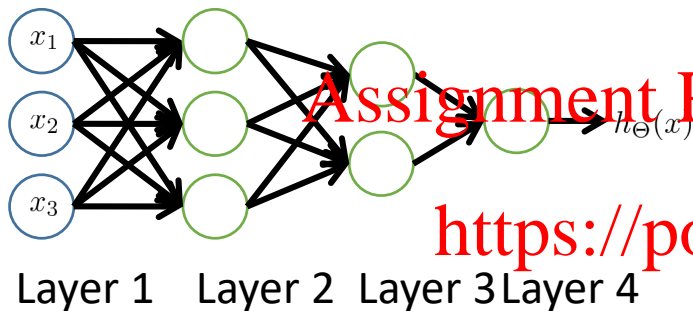
Assignment Project Exam Help

Network architectures

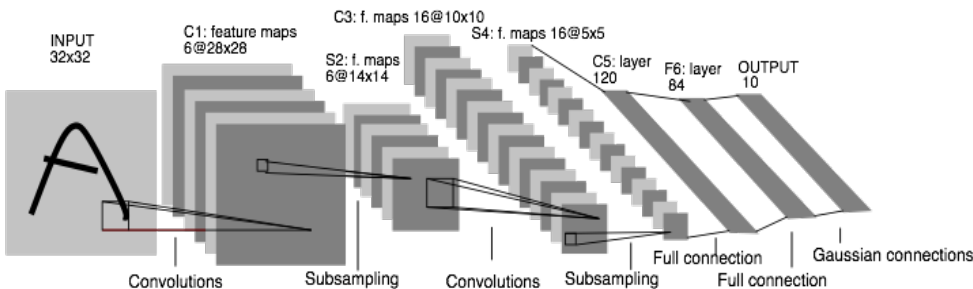
Add WeChat powcoder

Feed-forward

Fully connected

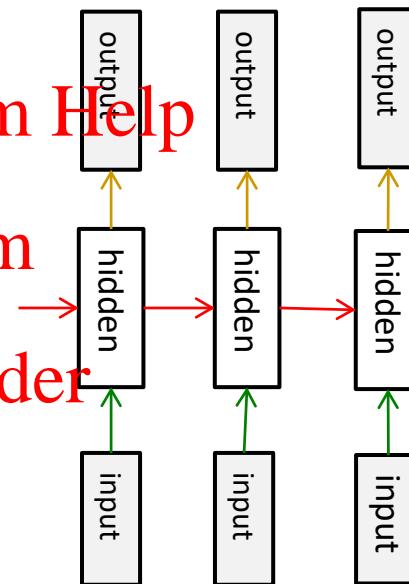


Convolutional



Recurrent

time →





Assignment Project Exam Help

Add WeChat powcoder

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Neural Networks IV

Recurrent Architectures

Assignment Project Exam Help

Recurrent Networks for Sequences of Data

Add WeChat powcoder

- Sequences in our world:

- Audio
- Text
- Video
- Weather
- Stock market

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



- Sequential data is why we build RNN architectures.
- RNNs are tools for making predictions about sequences.

Limitations of Feed-Fwd Networks

Add WeChat powcoder

- Limitations of feed-forward networks

Assignment Project Exam Help

- **Fixed length**

<https://powcoder.com>

Inputs and outputs are of fixed lengths

Add WeChat powcoder

- **Independence**

Data (example: images) are independent of one another

Assignment Project Exam Help

Advantages of RNN Models

Add WeChat powcoder

- What feed-forward networks cannot do

Assignment Project Exam Help

- **Variable length**

<https://powcoder.com>
“We would like to accommodate temporal sequences of various lengths.”

Add WeChat powcoder

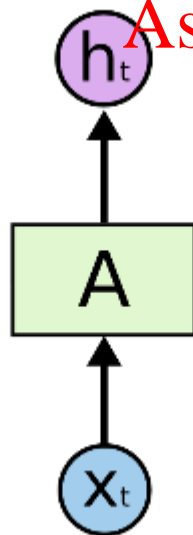
- **Temporal dependence**

“To predict where a pedestrian is at the next point in time, this depends on where he/she were in the previous time step.”

Assignment Project Exam Help

Vanilla Neural Network (NN)

Add WeChat powcoder

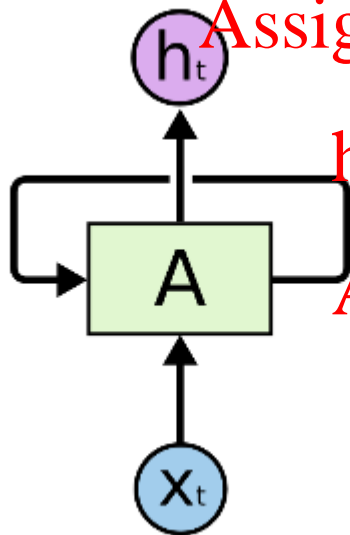


- NN
- x_t : input/event
 h_t : output/prediction
A : chunk of NN

Every input is treated independently.

Recurrent Neural Network (RNN)

Add WeChat powcoder



Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

- RNN

The loop allows information to be passed from one time step to the next.

Now we are modeling the dynamics.

Recurrent Neural Network (RNN)

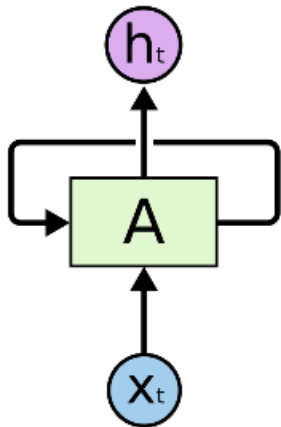
Add WeChat powcoder

- A recurrent neural network can be thought of as multiple copies of the same network, each passing a message to a successor.

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



Recurrent Neural Network (RNN)

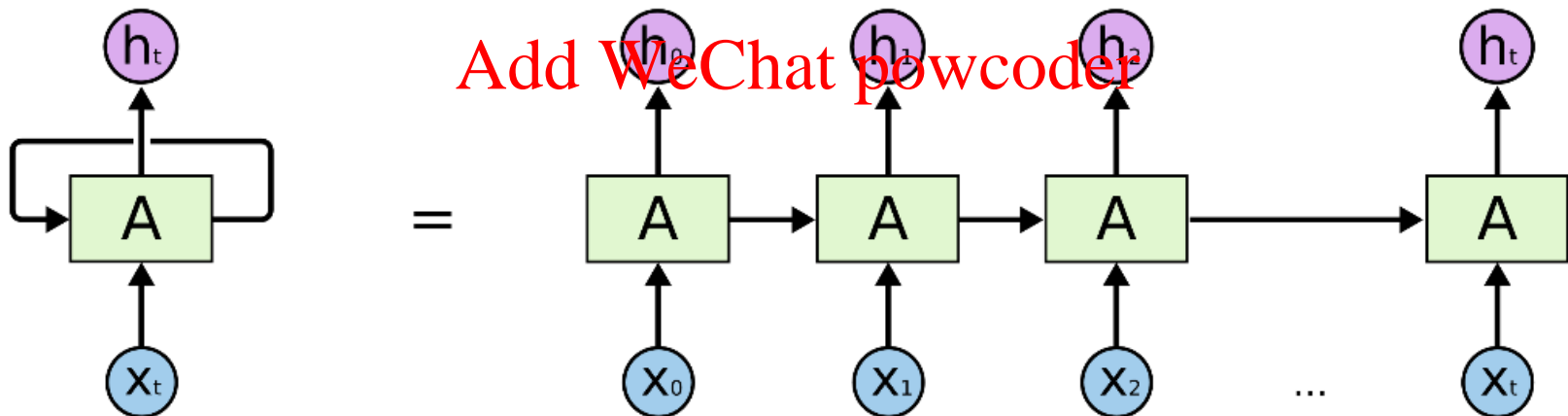
Add WeChat powcoder

- A recurrent neural network can be thought of as multiple copies of the same network, each passing a message to a successor.

Assignment Project Exam Help

<https://powcoder.com>

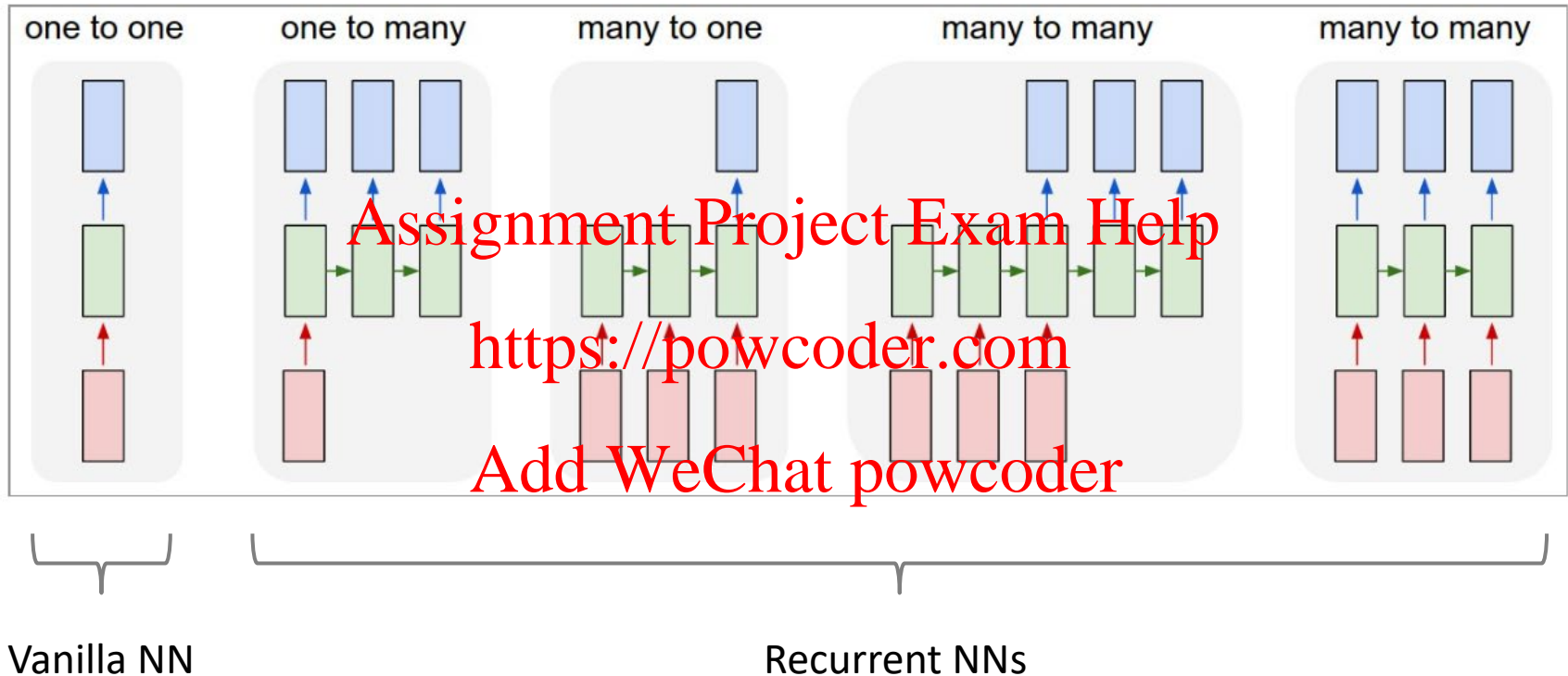
Add WeChat powcoder



Assignment Project Exam Help

RNN Architectures

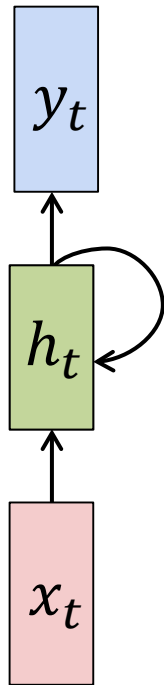
Add WeChat powcoder



Assignment Project Exam Help

Recurrent Neural Network

The state consists of a single "hidden" vector h .



Assignment Project Exam Help

$$h_t = f_W(h_{t-1}, x_t)$$

<https://powcoder.com>

$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t)$$

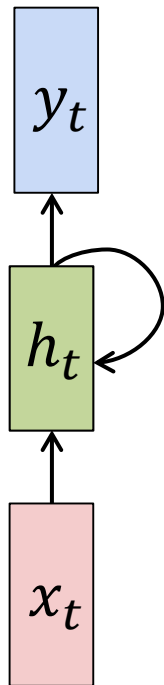
$$y_t = W_{hy}h_t$$

parameters

Assignment Project Exam Help

Recurrent Neural Network

The state consists of a single "hidden" vector \mathbf{h} .



Assignment Project Exam Help

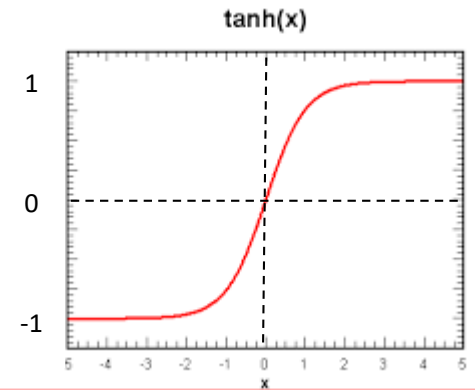
$$h_t = f_W(h_{t-1}, x_t)$$

<https://powcoder.com>

$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t)$$

$$y_t = W_{hy}h_t$$

activation function
(elementwise)



$$g(x) = 2\sigma(2x) - 1$$



Assignment Project Exam Help

Add WeChat powcoder

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Neural Networks IV

Example: Character RNN

Assignment Project Exam Help

Add WeChat powcoder

**Character-level
language model
example**

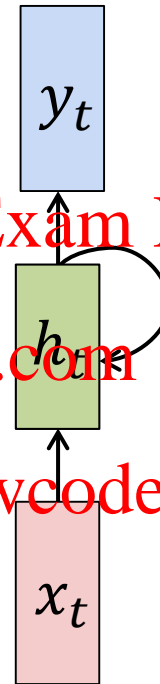
Vocabulary:
[h,e,l,o]

Example training
sequence:
“hello”

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



Assignment Project Exam Help

Add WeChat powcoder

**Character-level
language model
example**

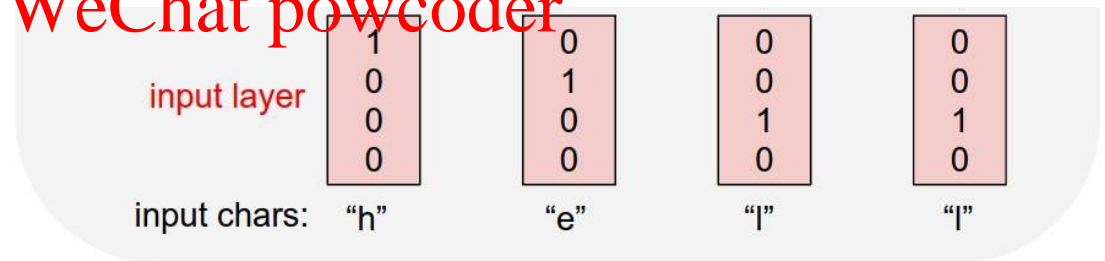
Vocabulary:
[h,e,l,o]

Example training
sequence:
“hello”

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



Assignment Project Exam Help

Add WeChat powcoder

Character-level
language model
example

$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t)$$

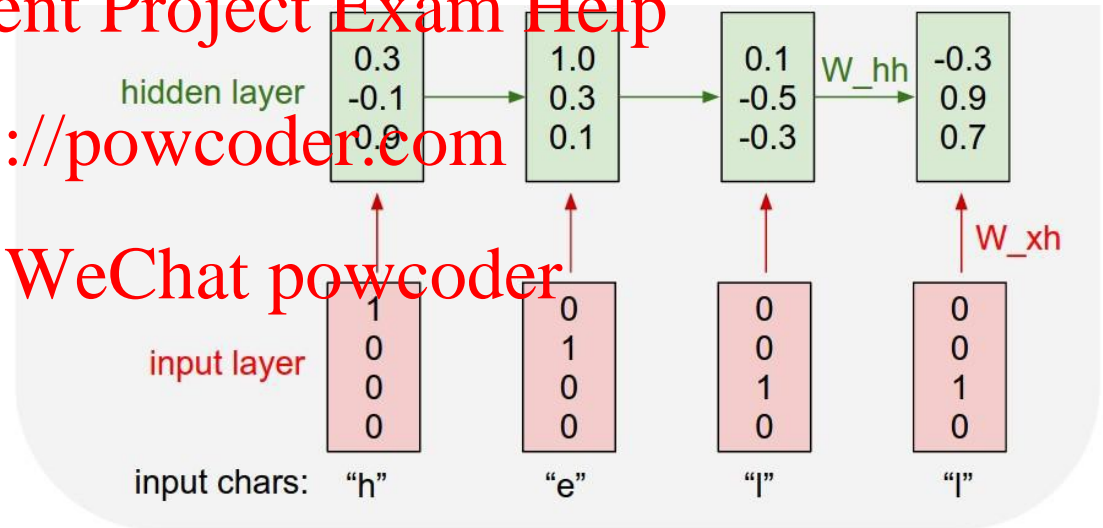
Vocabulary:
[h,e,l,o]

Example training
sequence:
“hello”

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



Assignment Project Exam Help

Add WeChat powcoder

**Character-level
language model
example**

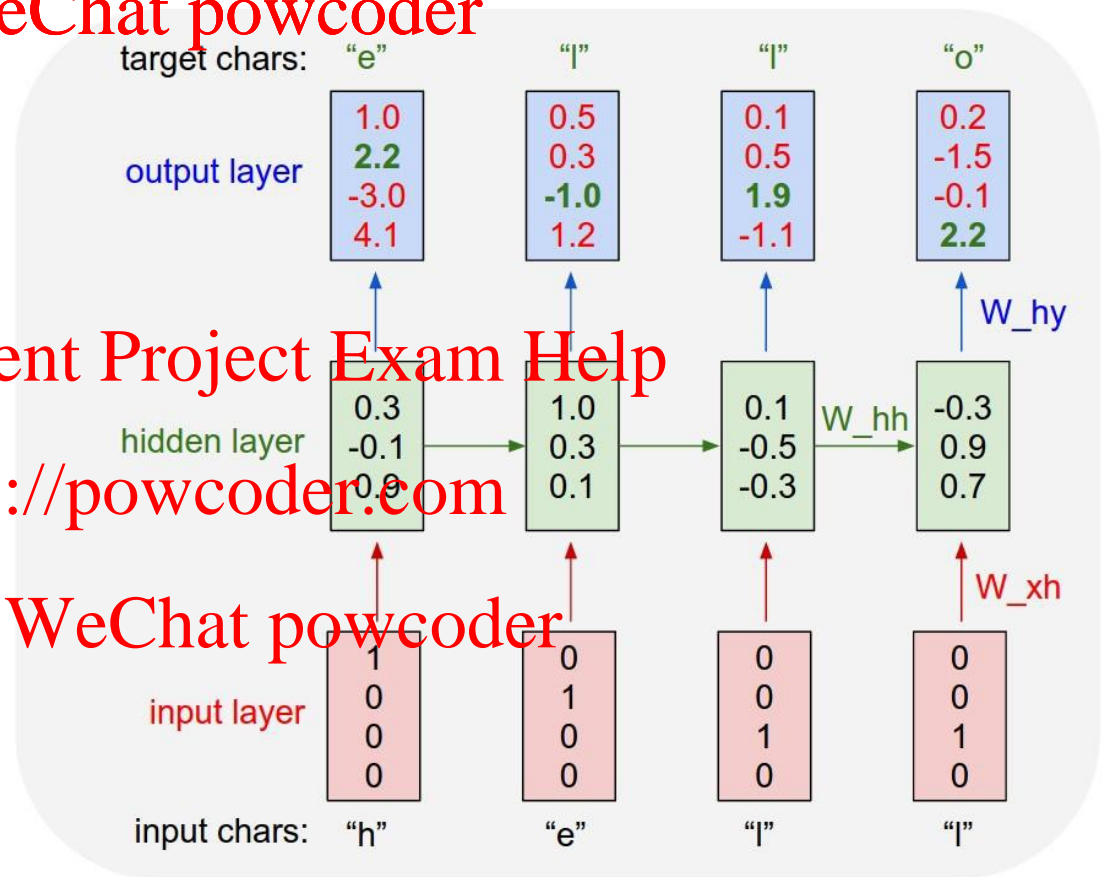
Vocabulary:
[h,e,l,o]

Example training
sequence:
“hello”

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



Assignment Project Exam Help

[min-char-rnn.py](#) gist: 112 lines of Python

Add WeChat powcoder

```
1 """
2 Minimal character-level Vanilla RNN model, written by Andrej Karpathy (aka @karpathy)
3 BSD license
4 """
5 import numpy as np
6
7 # data I/O
8 data = open('input.txt', 'r').read() # should be simple plain text file
9 chars = list(set(data))
10 data_size, vocab_size = len(data), len(chars)
11 print 'data has %d characters, %d unique.' % (data_size, vocab_size)
12 char_to_ix = { ch:i for i,ch in enumerate(chars) }
13 ix_to_char = { ix:i for i, ch in enumerate(chars) }
14
15 # hyperparameters
16 hidden_size = 100 # size of hidden layer of neurons
17 seq_length = 25 # number of steps to unroll the RNN for
18 learning_rate = 1e-1
19
20 # model parameters
21 wxh = np.random.randn(hidden_size, vocab_size)*0.01 # input to hidden
22 whh = np.random.randn(hidden_size, hidden_size)*0.01 # hidden to hidden
23 why = np.random.randn(hidden_size, vocab_size)*0.01 # hidden to output
24 bh = np.zeros((hidden_size, 1)) # hidden bias
25 by = np.zeros((vocab_size, 1)) # output bias
26
27 def lossFun(inputs, targets, hprev):
28     """
29     inputs, targets are both list of integers
30     hprev is Nx1 array of initial hidden state
31     returns the loss, gradients on model parameters, and last hidden state
32     """
33     xs, hs, ys, ps = {}, {}, {}, {}
34     hs[-1] = np.copy(hprev)
35     loss = 0
36     # forward pass
37     for t in xrange(len(inputs)):
38         xs[t] = np.zeros((vocab_size,1)) # encode in 1-hot representation
39         xs[t][inputs[t]] = 1
40         hs[t] = np.tanh(np.dot(wxh, xs[t]) + np.dot(whh, hs[t-1]) + bh) # hidden state
41         ys[t] = np.dot(why, hs[t]) + by # unnormalized log probabilities for next chars
42         ps[t] = np.exp(ys[t]) / np.sum(np.exp(ys[t])) # probabilities for next chars
43         loss += -np.log(ps[t][targets[t],0]) # softmax (cross-entropy loss)
44     # backward pass: compute gradients going backwards
45     dwhx, dwhh, dwhy = np.zeros_like(wxh), np.zeros_like(whh), np.zeros_like(why)
46     dbh, dby = np.zeros_like(bh), np.zeros_like(by)
47     dhnext = np.zeros_like(hs[0])
48     for t in reversed(xrange(len(inputs))):
49         dy = np.copy(ps[t])
50         dy[targets[t]] -= 1 # backprop into y
51         dwhy += np.dot(dy, hs[t].T)
52         dby += dy
53         dh = np.dot(why.T, dy) + dhnext # backprop into h
54         ddraw = (1 - hs[t]**2) * dh # backprop through tanh nonlinearity
55         dbh += ddraw
56         dwhx += np.dot(ddraw, xs[t].T)
57         dwhh += np.dot(ddraw, hs[t-1].T)
58         dhnext = np.dot(whh.T, ddraw)
59     for dparam in [dwhx, dwhh, dwhy, dbh, dby]:
60         np.clip(dparam, -5, 5, out=dparam) # clip to mitigate exploding gradients
61     return loss, dwhx, dwhh, dwhy, dbh, dby, hs[len(inputs)-1]
```

```
62
63 def sample(h, seed_ix):
64     """
65     sample a sequence of integers from the model
66     h is memory state, seed_ix is seed letter for first time step
67     """
68     x = np.zeros((vocab_size, 1))
69     x[seed_ix] = 1
70     ixes = []
71     for t in xrange(n):
72         h = np.tanh(np.dot(wxh, x) + np.dot(whh, h) + bh)
73         y = np.dot(why, h) + by
74         p = np.exp(y) / np.sum(np.exp(y))
75         ix = np.random.choice(range(vocab_size), p=p.ravel())
76         x = np.zeros((vocab_size, 1))
77         x[ix] = 1
78         ixes.append(ix)
79     return ixes
80
81 n, p = 0, 0
82 mwhx, mwhh, mwhy = np.zeros_like(wxh), np.zeros_like(whh), np.zeros_like(why)
83 mdbh, mwhy = np.zeros_like(bh), np.zeros_like(by) # memory variables for Adagrad
84 smooth_loss = np.zeros((vocab_size, seq_length)) # smooth loss over n iterations
85 while True:
86     # prepare inputs (we're sweeping from left to right in steps seq_length long)
87     if p+seq_length+1 >= len(data) or n == 0:
88         hprev = np.zeros((hidden_size,1)) # reset RNN memory
89         p = 0 # go from start of data
90         inputs = [char_to_ix[ch] for ch in data[p:p+seq_length]]
91         targets = [ch_to_ix[ch] for ch in data[p+1:p+seq_length+1]]
92
93     # sample from the model now and then
94     if n % 100 == 0:
95         sample_ix = sample(hprev, inputs[0], 200)
96         txt = ''.join(ix_to_char[ix] for ix in sample_ix)
97         print '----\n %s \n----' % (txt, )
98
99     # forward sequence characters through the net and fetch gradient
100     loss, dwhx, dwhh, dwhy, dbh, dby, hprev = lossFun(inputs, targets, hprev)
101     smooth_loss = smooth_loss * 0.999 + loss * 0.001
102     if n % 100 == 0: print 'iter %d, loss: %f' % (n, smooth_loss) # print progress
103
104     # perform parameter update with Adagrad
105     for param, dparam, mem in zip([wxh, whh, why, bh, by],
106                                   [dwhx, dwhh, dwhy, dbh, dby],
107                                   [mwhx, mwhh, mwhy, mdbh, mwhy]):
108         mem += dparam * dparam
109         param += -learning_rate * dparam / np.sqrt(mem + 1e-8) # adagrad update
110
111     p += seq_length # move data pointer
112     n += 1 # iteration counter
```

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

(<https://gist.github.com/karpathy/d4dee566867f8291f086>)

Assignment Project Exam Help

Training text

Add WeChat powcoder

<p class="clear">> Products: Laser Printers The fundamental everyday requirement for mono and colour laser printing throughout today's office is perfectly met with the extensive Epson laser printer range. The latest AcuLaser printer range offers users exceptionally

Epson AcuLaser C1900 Networked compact colour laser printer for professional enterprises. Businesses have been denied simple and affordable colour laser printing for far too long. The traditionally high costs and poor speeds of colour lasers has left many offices looking a bit, well, grey. But not any more with the Epson-AcuLaser C1900. Epson brings both colour and monochrome laser printing together at a black and white price. more Where to Buy Support Epson AcuLaser C2000 The fastest colour laser printer in its class. The perfect printer for small businesses and work groups, the Epson-AcuLaser C2000 prints high volumes in black and white and vibrant colour, at high speed and with low running costs. more Where to Buy High quality resolution: 2400dpi RIT* Large paper capacity: 600 sheets, expandable up to 1,600 sheets. Compatible Windows and Mac. High speed USB and Ethernet! 10/100 Base Tx Ethernet interfaces as standard* *Epson AcuLaser Resolution Improvement Technology* Epsonnet 10/100 Base Tx Ethernet standard with Epson AcuLaser C2000i model only. AcuLaser C1900 64MB Memory, 100 sheet MP Tray, 500 sheet cassette, Duplex printing as standard AcuLaser C2000i 64MB Memory, 100 sheet MP Tray, 500 sheet cassette, Duplex printing as standard Ethernet Interface Networked compact colour laser printer for professional enterprises. Businesses have been denied simple and affordable colour laser printing for far too long. The traditionally high costs and poor speeds of colour lasers has left many offices looking a bit, well, grey. But not any more with the Epson-AcuLaser C2000. Epson brings both colour and monochrome laser printing together at a black and white price. Key features cost effective mono printing for day to day business needs and vivid versatile colour when required. search Search Epson UK Epson AcuLaser C500 Outstanding professional colour printing for business. Add colour to your business with the Epson AcuLaser C500 from Epson. Its perfect for the smaller workgroup, being a compact and cost effective laser printing workhorse that offers amazing colour output as well as high performance black and white production. more Where to Buy Support As cost efficient to run as a mono-only laser printer. Paper capacity of 700 sheets from two media sources. Easy to operate with advanced print driver. Memory expandable from 32MB to 1024MB. Pre-configured models available with Wireless 802.11b, Adobe® PostScript® 3 Level 3™ and two-sided printing. The AcuLaser C1900 is available in 5 configurations. - AcuLaser C1900S: with 32MB, 200 Sheet MP Tray, 10/100BaseTX Networking - AcuLaser C1900: with 32MB, 200 Sheet MP Tray, 500 Sheet Cassette, 10/100BaseTX Networking - Support Epson AcuLaser C4100 High performance colour lasers for all your business printing needs. The Epson AcuLaser C4100 provides businesses with a high performance colour and monochrome printing solution. A solid crucial colour by your business, while producing high quality monochrome and colour solutions are available in one. more Where to Buy Support Epson AcuLaser C5600 A3W colour laser printer. Epson AcuLaser C5600 is the perfect professional printing solution. It's designed to deliver exceptional high colour and mono output on a range of media formats from C5 up to A3W in size. The Epson AcuLaser C5600 is able to achieve superb print quality by utilising a combination of Epson's exclusive AcuLaser Colour Laser Technologies. more Where to Buy Support AcuLaser C1900PS: with Adobe® PostScript® 3™, 96MB, 200 Sheet MP Tray, 500 Sheet Cassette, 10/100BaseTX Networking - AcuLaser C1900D: with Duplex unit (two sided printing) 96MB, 200 Sheet MP Tray, 500 Sheet Cassette, 10/100BaseTX Networking - AcuLaser C1900 WFi: with 32MB, 200 Sheet MP Tray, 500 Sheet Cassette, Wireless Networking facility. Add colour to your business with the Epson AcuLaser C800 from Epson. Its perfect for the smaller workgroup, being a compact and cost effective laser printing workhorse that offers amazing colour output as well as high. Support Epson AcuLaser C4000 High performance colour laser. The Epson AcuLaser C4000 provides businesses with high performance colour and monochrome printing solutions. more Where to Buy Support Epson AcuLaser C8100 High speed A3 colour laser printer. With separate black and white and colour print engines when you can have the Epson AcuLaser C8100! Epson has taken the lead in laser technology to deliver a complete high-performance solution for your colour and mono printing needs. Support EPL-6200L High performance A4 mono professional printer. The Epson EPL-6200L is a professional performance quickly, easily, reliably and cost-effectively, and are perfect for users who need high levels of productivity at a low investment. more Where to Buy Support EPL-6200 High performance A4 mono professional printer. The Epson EPL-6200 and EPL-6200L are the ideal printing solutions for small to medium workgroups and professional users. They deliver professional performance quickly, easily, reliably and cost-effectively, and are perfect for users who need high levels of laser quality and productivity at a low investment, more performance black and white production. For the first time, you can now bring the power of high quality colour to your documents without suffering the high costs or low speeds traditionally associated with colour

y

RNN

x

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Assignment Project Exam Help

Add WeChat powcoder

Sonnet 116 – Let me not ...

by William Shakespeare

Let me not to the marriage of true minds
Admit impediments. Love is not love
Which alters when it alteration finds,
Or bends with the remover to remove:
O no! it is an ever-fixed mark
That looks on tempests and is never shaken;
It is the star so even in our worst
Whose worth's unknown, although his height be taken.
Love's not Time's fool, though rosy lips and cheeks
Within his bending sickle's compass come;
Love alters not with his brief hours and weeks,
But bears it out even to the edge of doom.
If this be error and upon me proved,
I never writ, nor no man ever loved.

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Assignment Project Exam Help

Add WeChat powcoder

at first:

tyntd-iafhatawiaoihrdemot lytdws e ,tfti, astai f ogoh eoase rrranbyne 'nhtnee e
plia tklrqd t o idoe ns,smtt h ne etie h,hregtrs nigtiike,aoaenns lng

train more

"Tmont thithey" fomesscerliund
Keushey. Thom here
sheulke, anmerenith ol sivh I lalterthend Bleipile shuw y fil on aseterlome
coaniogenn. "Se iin thond hon at Mel no totion in ther tize."

train more

Aftair fall unsuch that the hall for Prince Velzonski's that me of
her hearly, and behs to so arwage fiving were to it beloge, pavu say falling misfort
how, and Gogition is so overelal and ofar.

train more

"Why do what that day," replied Natasha, and wishing to himself the fact the
princess, Princess Mary was easier, fed in had oftened him.
Pierre aking his soul came to the packs and drove up his father-in-law women.

Assignment Project Exam Help

Add WeChat powcoder

PANDARUS:

Alas, I think he shall be come approached and the day
When little strain would be attain'd into being never fed,
And who is but a chain and subjects of his death,
I should not sleep.

Second Senator:

They are away this miseries, produced upon my soul,
Breaking and strongly should be buried, when I perish
The earth and thoughts of man, states.

DUKE VINCENTIO:

Well, your wit is in the care of side and that

Second Lord:

They would be ruled after this chamber, and
my fair nudes begun out of the fact, to be conveyed
Whose noble souls I'll have the heart of the wars.

Clown:

Come, sir, I will make did behold your worship.

VIOLA:

I'll drink it.

VIOLA:

Why, Salisbury must find his flesh and thought
That which I am not aps, not a man and in fire,
To show the reining of the raven and the wars
To grace my hand reproach within, and not a fair are hand,
That Caesar and my goodly father's world;
When I was heaven of presence and our fleets,
We spare with hours, but cut thy council I am great,
Murdered and by thy master's ready there
My power to give thee but so much as hell:
Some service in the noble bondman here,
Would show him to her wine.

KING LEAR:

O, if you were a feeble sight, the courtesy of your law,
Your sight and several breath, will wear the gods
With his heads, and my hands are wonder'd at the deeds,
So drop upon your lordship's head, and your opinion
Shall be against your honour.

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Assignment Project Exam Help

Add WeChat powcoder

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



Neural Networks IV

Learning in RNNs

Assignment Project Exam Help

Forward pass

Add WeChat powcoder

- Forward pass through time

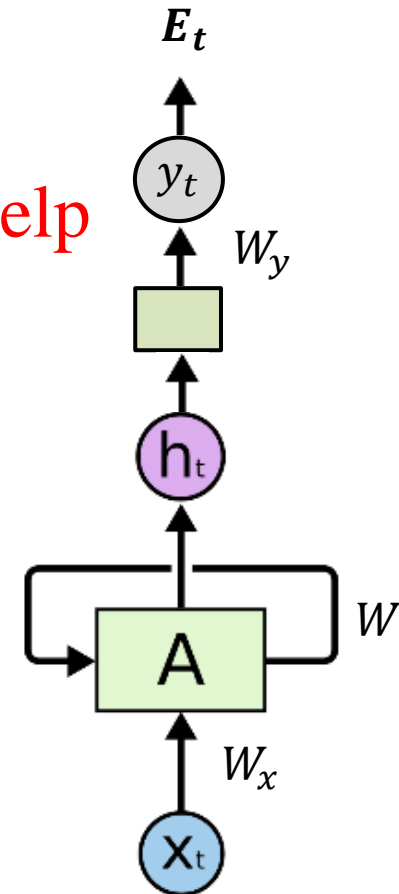
Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

$$h_t = W\phi(h_{t-1}) + W_x x_t$$

$$y_t = W_y \phi(h_t)$$



Assignment Project Exam Help

Recurrent Neural Network (RNN)

Add WeChat powcoder

Assignment Project Exam Help

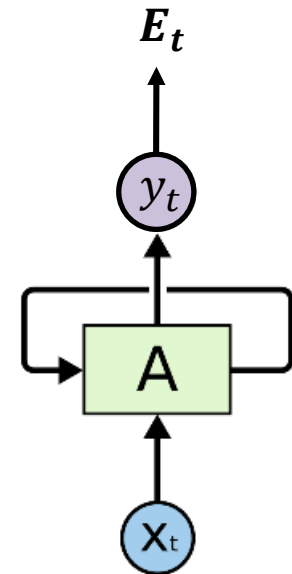
<https://powcoder.com>

Add WeChat powcoder

Aside: Forward pass

$$h_t = W\phi(h_{t-1}) + W_x x_t$$

$$y_t = W_y \phi(h_t)$$



Assignment Project Exam Help

Recurrent Neural Network (RNN)

Add WeChat powcoder

- Error or cost is computed for each prediction.

Aside: Forward pass

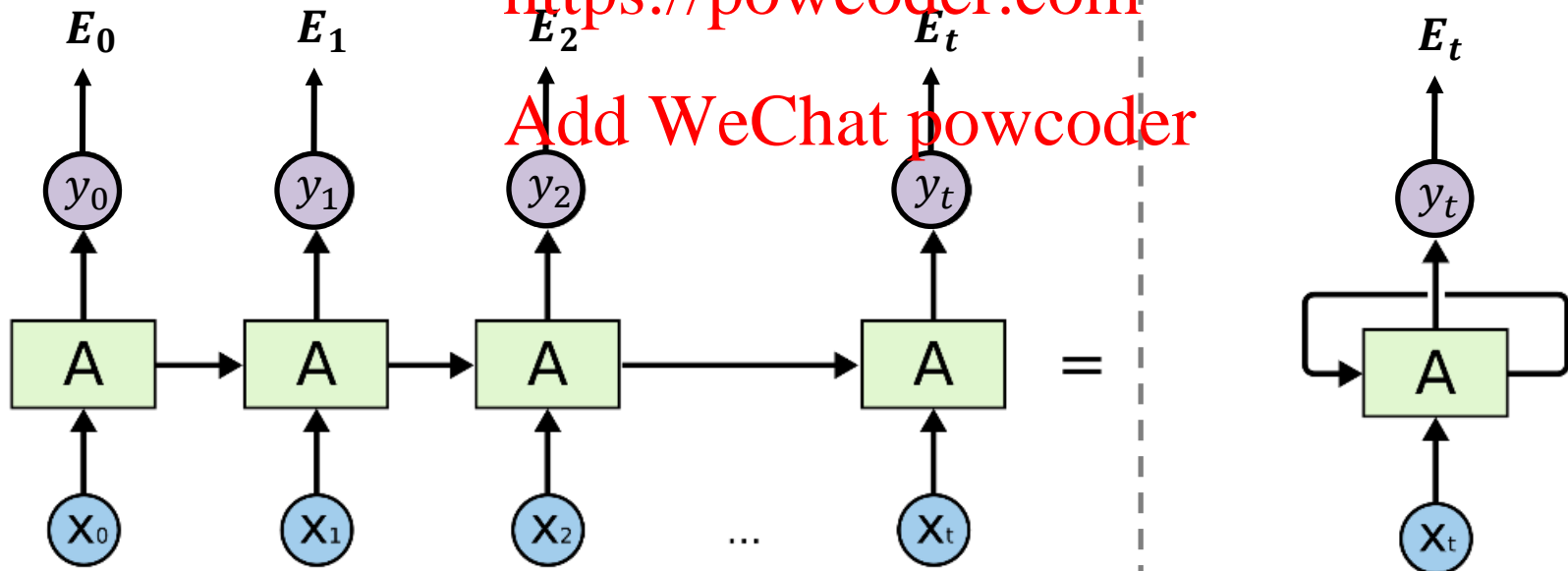
$$h_t = W\phi(h_{t-1}) + W_x x_t$$

$$y_t = W_y \phi(h_t)$$

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder



Assignment Project Exam Help

Backprop Through Time

Add WeChat powcoder

- Backpropagation through time

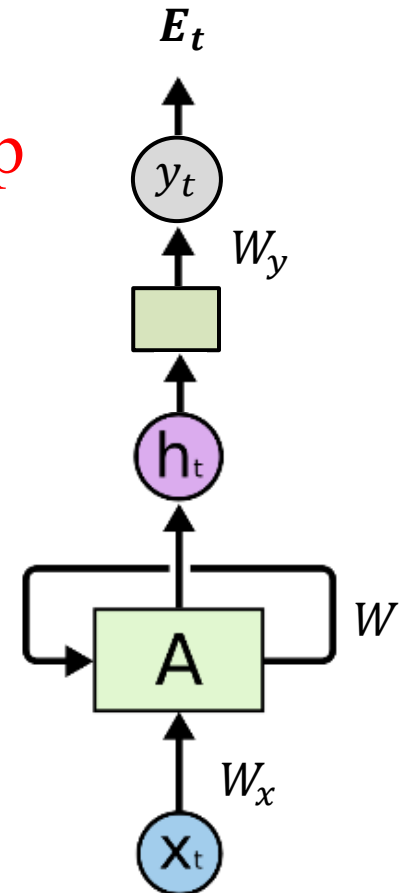
$$\frac{\partial E}{\partial W} = \sum_{t=1}^T \frac{\partial E_t}{\partial W}$$

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Aside: Forward pass



Assignment Project Exam Help

BP TT

Add WeChat powcoder

- Backpropagation through time

$$\frac{\partial E}{\partial W} = \sum_{t=1}^T \boxed{\frac{\partial E_t}{\partial W}}$$

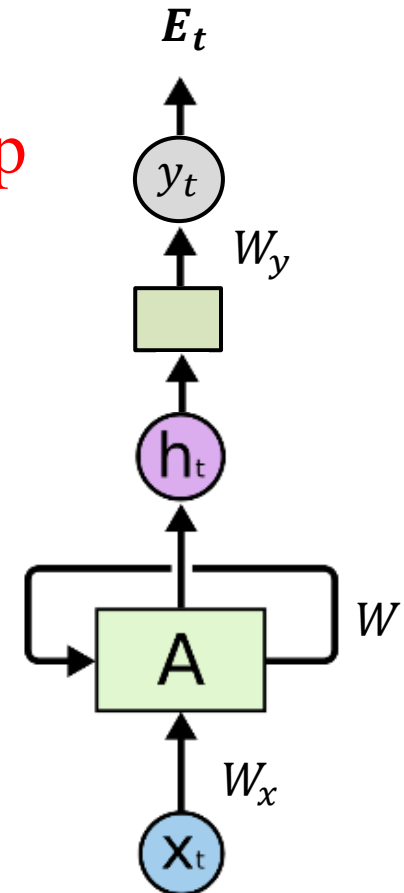
Assignment Project Exam Help

<https://powcoder.com>

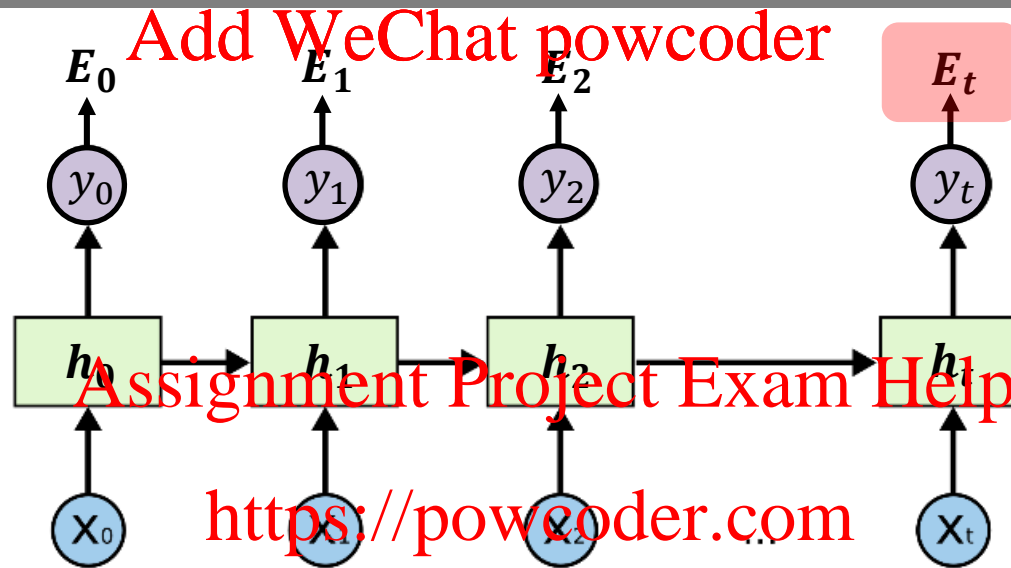
Add WeChat powcoder

$$\frac{\partial E_t}{\partial W} = \sum_{k=1}^t \frac{\partial E_t}{\partial y_t} \frac{\partial y_t}{\partial h_t} \frac{\partial h_t}{\partial h_k} \frac{\partial h_k}{\partial W}$$

Aside: Forward pass



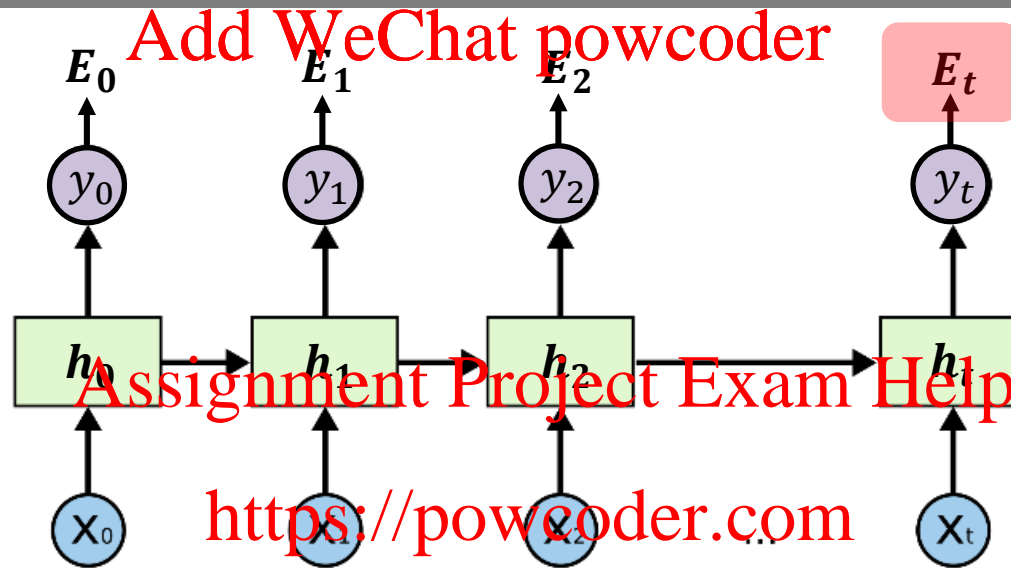
Assignment Project Exam Help



Add WeChat powcoder

$$\frac{\partial E_t}{\partial W} = \sum_{k=1}^t \frac{\partial E_t}{\partial y_t} \frac{\partial y_t}{\partial h_t} \frac{\partial h_t}{\partial h_k} \frac{\partial h_k}{\partial W}$$

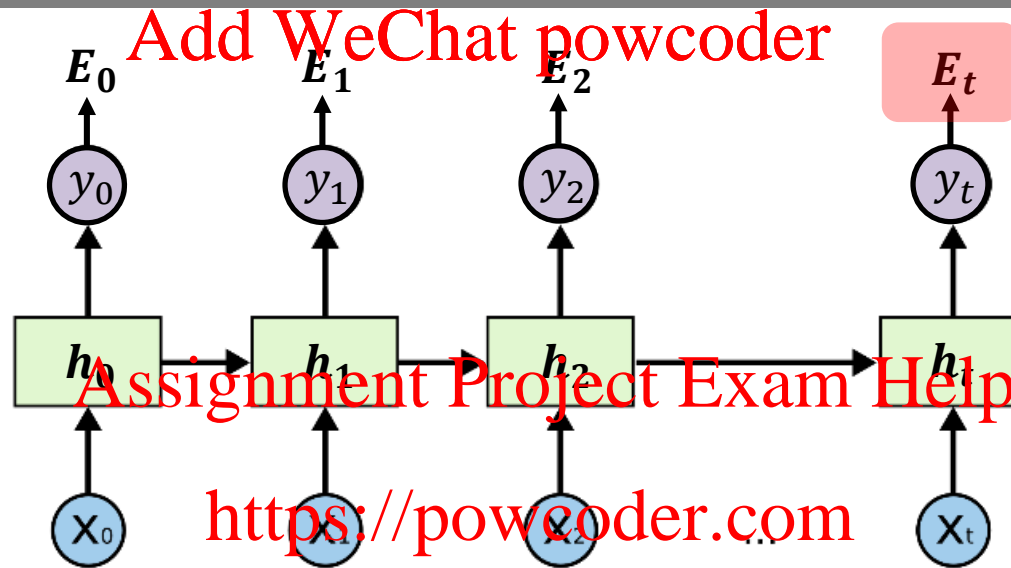
Assignment Project Exam Help



Add WeChat powcoder

$$\frac{\partial E_t}{\partial W} = \sum_{k=1}^t \frac{\partial E_t}{\partial y_t} \frac{\partial y_t}{\partial h_t} \frac{\partial h_t}{\partial h_k} \frac{\partial h_k}{\partial W}$$

Assignment Project Exam Help

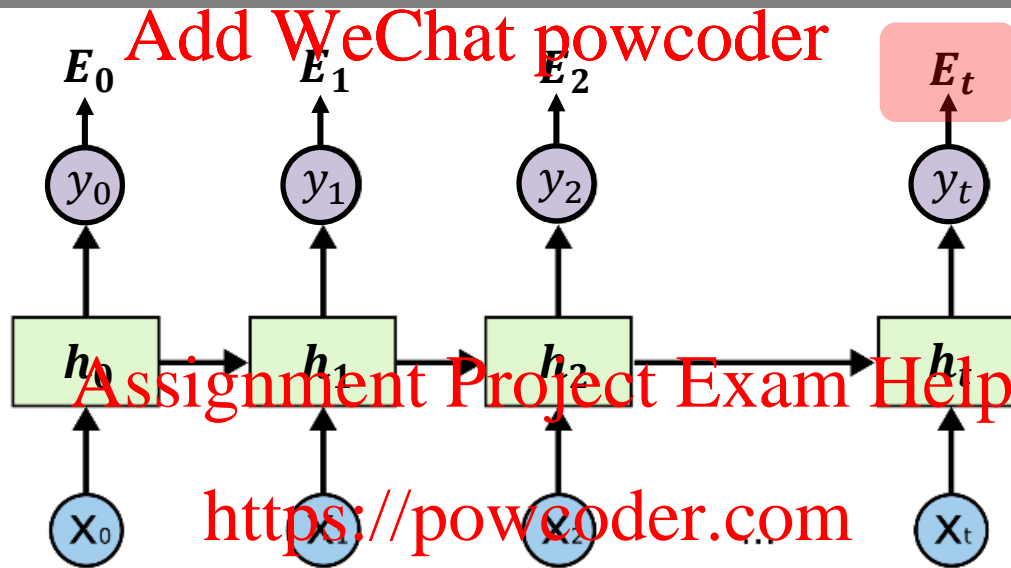


Add WeChat powcoder

$$\frac{\partial E_t}{\partial W} = \sum_{k=1}^t \frac{\partial E_t}{\partial y_t} \frac{\partial y_t}{\partial h_t} \frac{\partial h_t}{\partial h_k} \frac{\partial h_k}{\partial W}$$

$$\frac{\partial h_t}{\partial h_k} = \prod_{i=k+1}^t \frac{\partial h_i}{\partial h_{i-1}}$$

Assignment Project Exam Help



Add WeChat powcoder

$$\frac{\partial E_t}{\partial W} = \sum_{k=1}^t \frac{\partial E_t}{\partial y_t} \frac{\partial y_t}{\partial h_t} \frac{\partial h_t}{\partial h_k} \frac{\partial h_k}{\partial W}$$

$$\frac{\partial h_t}{\partial h_k} = \prod_{i=k+1}^t \frac{\partial h_i}{\partial h_{i-1}}$$

For example @ $t = 2$,

$$\frac{\partial h_2}{\partial h_0} = \prod_{i=1}^2 \frac{\partial h_i}{\partial h_{i-1}} = \frac{\partial h_1}{\partial h_0} \frac{\partial h_2}{\partial h_1}$$

Assignment Project Exam Help

Vanishing (and Exploding) Gradients

Add WeChat powcoder

$$\frac{\partial E_t}{\partial W} = \sum_{k=1}^t \frac{\partial E_t}{\partial y_t} \frac{\partial y_t}{\partial h_t} \frac{\partial h_t}{\partial h_k} \frac{\partial h_k}{\partial W}$$

Assignment Project Exam Help

<https://powcoder.com>

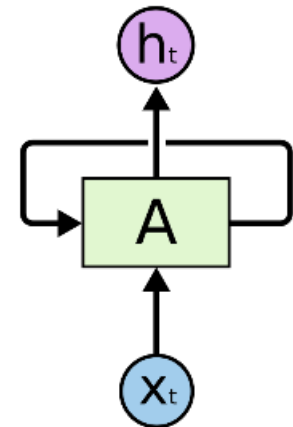
Add WeChat powcoder

$$\frac{\partial h_t}{\partial h_k} = \prod_{i=k+1}^t \frac{\partial h_i}{\partial h_{i-1}}$$

Aside: Forward pass

$$h_t = W\phi(h_{t-1}) + W_x x_t$$

$$y_t = W_y \phi(h_t)$$



Assignment Project Exam Help

Vanishing (and Exploding) Gradients

Add WeChat powcoder

$$\frac{\partial E_t}{\partial W} = \sum_{k=1}^t \frac{\partial E_t}{\partial y_t} \frac{\partial y_t}{\partial h_t} \frac{\partial h_t}{\partial h_k} \frac{\partial h_k}{\partial W}$$

Assignment Project Exam Help

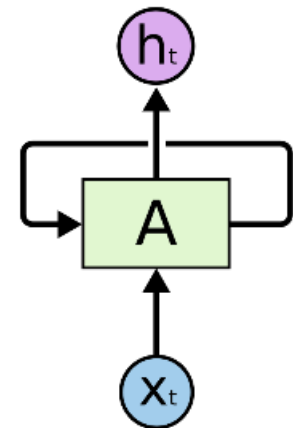
$$\frac{\partial h_t}{\partial h_k} = \prod_{i=k+1}^t \frac{\partial h_i}{\partial h_{i-1}} = \prod_{i=k+1}^t W^T \text{diag}[\phi'(h_{i-1})]$$

Add WeChat powcoder

Aside: Forward pass

$$h_t = W\phi(h_{t-1}) + W_x x_t$$

$$y_t = W_y \phi(h_t)$$



Assignment Project Exam Help

Vanishing (and Exploding) Gradients

Add WeChat powcoder

$$\frac{\partial E_t}{\partial W} = \sum_{k=1}^t \frac{\partial E_t}{\partial y_t} \frac{\partial y_t}{\partial h_t} \frac{\partial h_t}{\partial h_k} \frac{\partial h_k}{\partial W}$$

Assignment Project Exam Help

$$\frac{\partial h_t}{\partial h_k} = \prod_{i=k+1}^t \frac{\partial h_i}{\partial h_{i-1}} = \prod_{i=k+1}^t W^T \text{diag}[\phi'(h_{i-1})]$$

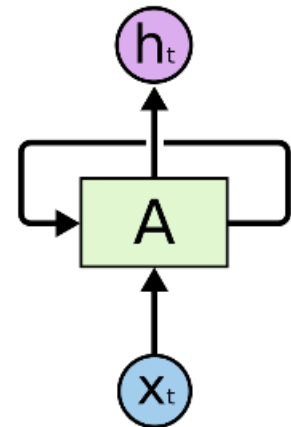
Add WeChat powcoder

$$\left\| \frac{\partial h_i}{\partial h_{i-1}} \right\|$$

Aside: Forward pass

$$h_t = W\phi(h_{t-1}) + W_x x_t$$

$$y_t = W_y \phi(h_t)$$



Assignment Project Exam Help

Vanishing (and Exploding) Gradients

Add WeChat powcoder

$$\frac{\partial E_t}{\partial W} = \sum_{k=1}^t \frac{\partial E_t}{\partial y_t} \frac{\partial y_t}{\partial h_t} \frac{\partial h_t}{\partial h_k} \frac{\partial h_k}{\partial W}$$

Assignment Project Exam Help

$$\frac{\partial h_t}{\partial h_k} = \prod_{i=k+1}^t \frac{\partial h_i}{\partial h_{i-1}} = \prod_{i=k+1}^t W^T \text{diag}[\phi'(h_{i-1})]$$

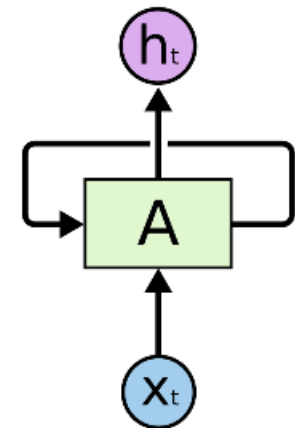
Add WeChat powcoder

$$\left\| \frac{\partial h_i}{\partial h_{i-1}} \right\| \leq \|W^T\| \|\text{diag}[\phi'(h_{i-1})]\|$$

Aside: Forward pass

$$h_t = W\phi(h_{t-1}) + W_x x_t$$

$$y_t = W_y \phi(h_t)$$



Assignment Project Exam Help

Vanishing (and Exploding) Gradients

Add WeChat powcoder

$$\frac{\partial E_t}{\partial W} = \sum_{k=1}^t \frac{\partial E_t}{\partial y_t} \frac{\partial y_t}{\partial h_t} \frac{\partial h_t}{\partial h_k} \frac{\partial h_k}{\partial W}$$

Assignment Project Exam Help

$$\frac{\partial h_t}{\partial h_k} = \prod_{i=k+1}^t \frac{\partial h_i}{\partial h_{i-1}} = \prod_{i=k+1}^t W^T \text{diag}[\phi'(h_{i-1})]$$

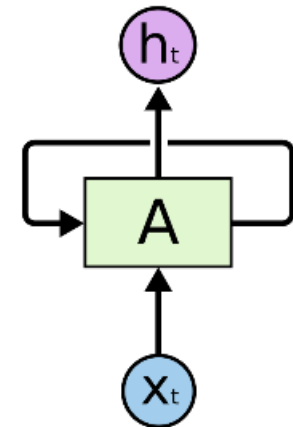
Add WeChat powcoder

$$\left\| \frac{\partial h_i}{\partial h_{i-1}} \right\| \leq \|W^T\| \|\text{diag}[\phi'(h_{i-1})]\| \leq \gamma_W \gamma_\phi$$

Aside: Forward pass

$$h_t = W\phi(h_{t-1}) + W_x x_t$$

$$y_t = W_y \phi(h_t)$$



Assignment Project Exam Help

Vanishing (and Exploding) Gradients

Add WeChat powcoder

$$\frac{\partial E_t}{\partial W} = \sum_{k=1}^t \frac{\partial E_t}{\partial y_t} \frac{\partial y_t}{\partial h_t} \frac{\partial h_t}{\partial h_k} \frac{\partial h_k}{\partial W}$$

Assignment Project Exam Help

$$\frac{\partial h_t}{\partial h_k} = \prod_{i=k+1}^t \frac{\partial h_i}{\partial h_{i-1}} = \prod_{i=k+1}^t W^T \text{diag}[\phi'(h_{i-1})]$$

Add WeChat powcoder

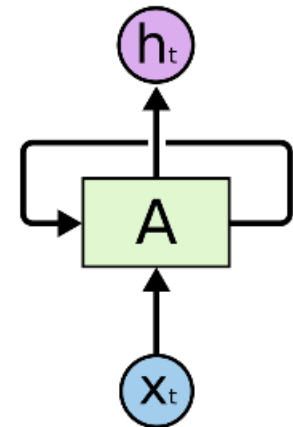
$$\left\| \frac{\partial h_i}{\partial h_{i-1}} \right\| \leq \|W^T\| \|\text{diag}[\phi'(h_{i-1})]\| \leq \gamma_W \gamma_\phi$$

$$\prod_{i=k+1}^t \left\| \frac{\partial h_i}{\partial h_{i-1}} \right\| \leq (\gamma_W \gamma_\phi)^{t-k}$$

Aside: Forward pass

$$h_t = W\phi(h_{t-1}) + W_x x_t$$

$$y_t = W_y \phi(h_t)$$



Assignment Project Exam Help

Vanishing (and Exploding) Gradients

Add WeChat powcoder

$$\frac{\partial E_t}{\partial W} = \sum_{k=1}^t \frac{\partial E_t}{\partial y_t} \frac{\partial y_t}{\partial h_t} \frac{\partial h_t}{\partial h_k} \frac{\partial h_k}{\partial W}$$

Assignment Project Exam Help

$$\frac{\partial h_t}{\partial h_k} = \prod_{i=k+1}^t \frac{\partial h_i}{\partial h_{i-1}} = \prod_{i=k+1}^t W^T \text{diag}[\phi'(h_{i-1})]$$

Add WeChat powcoder

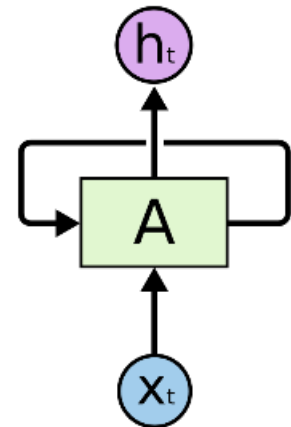
$$\left\| \frac{\partial h_i}{\partial h_{i-1}} \right\| \leq \|W^T\| \|\text{diag}[\phi'(h_{i-1})]\| \leq \gamma_W \gamma_\phi$$

$$\prod_{i=k+1}^t \left\| \frac{\partial h_i}{\partial h_{i-1}} \right\| \leq \underbrace{(\gamma_W \gamma_\phi)}_{\substack{<1 \text{ vanishing} \\ >1 \text{ exploding}}}^{t-k}$$

Aside: Forward pass

$$h_t = W\phi(h_{t-1}) + W_x x_t$$

$$y_t = W_y \phi(h_t)$$



Vanishing (and Exploding) Gradients

Add WeChat powcoder

- Exploding Gradients

- Easy to detect
- Clip the gradient at a threshold

Assignment Project Exam Help

<https://powcoder.com>

- Vanishing Gradients

Add WeChat powcoder

- More difficult to detect
- Architectures designed to combat the problem of vanishing gradients. Example: LSTMs by *Schmidhuber et al.*



Assignment Project Exam Help

Add WeChat powcoder

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder

Neural Networks IV

Training strategies

Assignment Project Exam Help

Universality

Add WeChat powcoder

- Why study neural networks in general?
 - Neural network can approximate any continuous function, even with a single hidden layer!
 - <http://neuralnetworksanddeeplearning.com/chap4.html>

<https://powcoder.com>

Add WeChat powcoder

Why Study Deep Networks?

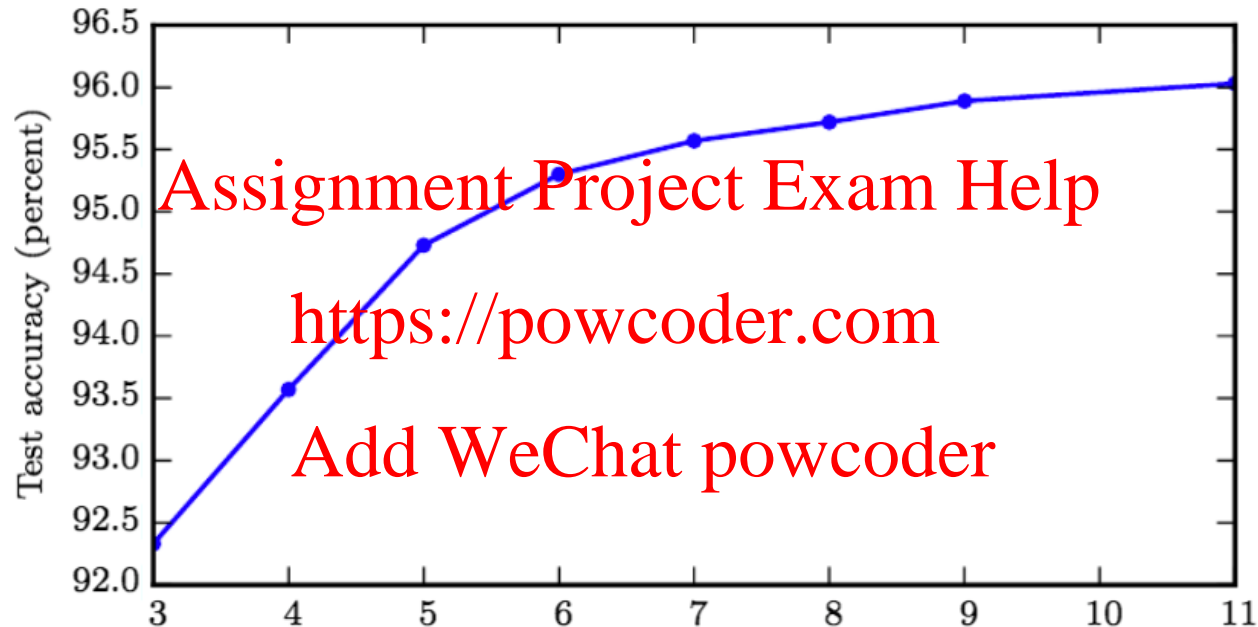
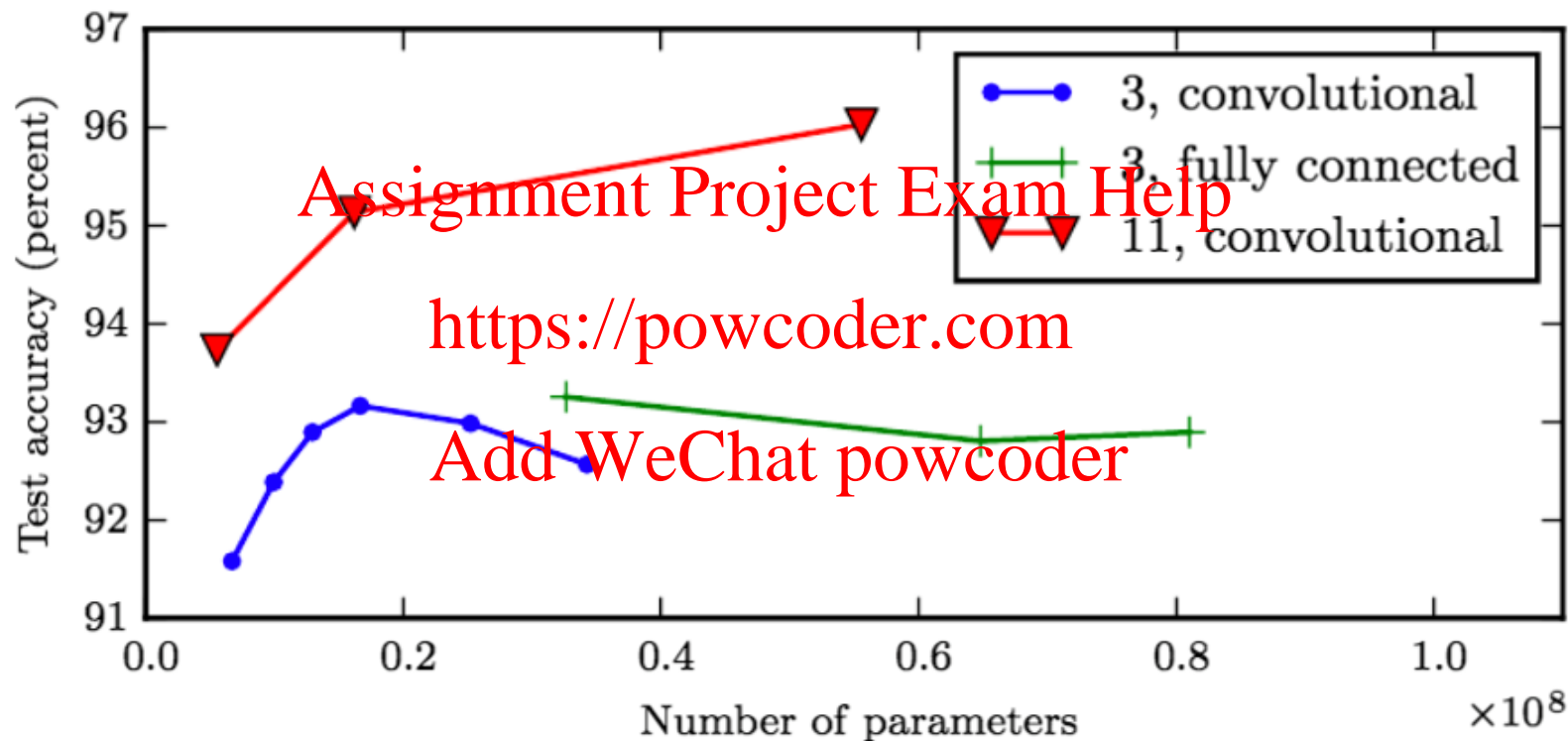


Figure 6.6: Empirical results showing that deeper networks generalize better when used to transcribe multi-digit numbers from photographs of addresses. Data from Goodfellow *et al.* (2014d). The test set accuracy consistently increases with increasing depth. See figure 6.7 for a control experiment demonstrating that other increases to the model size do not yield the same effect.

Assignment Project Exam Help

Efficiency of convnets

Add WeChat powcoder



But... Watch Out for Vanishing Gradients

- Consider a simple network, and perform backpropagation



- For simplicity, just a single neuron
- Sigmoid at every layer, $a_j = \sigma(w_j a_{j-1} + b_j)$, $a_j = \sigma(z_j)$
- Cost function C

- Gradient $\partial C / \partial b_1$ is a product of terms:

$$\partial C / \partial b_1 = \sigma'(z_1) w_2 \sigma'(z_2) w_3 \sigma'(z_3) w_4 \sigma'(z_4) (\partial C / \partial a_4)$$

Assignment Project Exam Help

Vanishing Gradients

Add WeChat powcoder

- Gradient of sigmoid is in $(0, 1/4)$
- Weights are also typically initialized in $(0, 1)$
- Products of small numbers \rightarrow small gradients
- Backprop does not change weights in earlier layers by much!
 - This is an issue with backprop, not with the model itself

RNNs: vanishing and exploding gradients

- Exploding: easy to fix, clip the gradient at a threshold
- Vanishing: More difficult to detect
- Architectures designed to combat the problem of vanishing gradients. Example: LSTMs by *Schmidhuber et al.*

Assignment Project Exam Help

Rectified Linear Units (RELU)

Add WeChat powcoder

- Alternative non-linearity:

$$g(x) = \max(0, x)$$

- Gradient of this function?
 - Note: need subgradient descent here.
- https://cs224d.stanford.edu/notebooks/vanishing_grad_example.html
- Increasing the number of layers can result in requiring exponentially fewer hidden units per layer (see “Understanding Deep Neural Networks with Rectified Linear Units”)
- Biological considerations
 - On some inputs, biological neurons have no activation
 - On some inputs, neurons have activation proportional to input

Assignment Project Exam Help

Other Activation Functions

Add WeChat powcoder

- Leaky ReLU: $g(x) = \max(0, x) + \alpha \min(0, x)$ ($\alpha \approx .01$)
- Tanh: $g(x) = 2\sigma(2x) - 1$
- Radial Basis Functions: $g(x) = \exp(-(w - x)^2 / \sigma^2)$
- Softplus: $g(x) = \log(1 + e^x)$
- Hard Tanh: $g(x) = \max(-1, \min(1, x))$
- Maxout: $g(x) = \max_{j \in \mathbb{G}} x_j$
-

Architecture Design and Training Issues

- How many layers? How many hidden units per layer? How to connect layers together? How to optimize?

- Cost functions
- L2/L1 regularization
- Data Set Augmentation
- Early Stopping
- Dropout
- Minibatch Training
- Momentum
- Initialization
- Batch Normalization

Assignment Project Exam Help

Next Class

Add WeChat powcoder

Computing cluster/Tensorflow Intro (next Thursday):

Intro to SCC and Tensorflow; please have laptops ready to follow along with the lecture. Expected to last 2 hours

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder