CS280 Fall 2018 Assignment 2
Part A

CNNs

Due in class, Nov 02, 2018

Name: MinJie 闵杰

Student ID: 10109867

1. Linear Regression(10 points)

- Linear regression has the form $E[y|x] = w_0 + \boldsymbol{w^T}x$. It is possible to solve for $\boldsymbol{w}$ and w_0 seperately. Show that

$$w_0 = \frac{1}{n}\sum_i y_i - \frac{1}{n}\sum_i x_i^T \boldsymbol{w} = \overline{y} - \overline{x}^T \boldsymbol{w}$$

- Show how to cast the problem of linear regression with respect to the absolute value loss function, $l(h, x, y) = |h(x) - y|$, as a linear program.

I 1.1 Solution:

$loss = \sum_{i=1}^{n}(y_i - w_0 - \boldsymbol{w^T x})$

To get w0, $\frac{\partial loss}{\partial w_0} = 2(n * w_0 - \sum_{i=1}^{n}(y_i - \boldsymbol{w^T x_i}))$

Let $\frac{\partial loss}{\partial w_0} = 0$, then $w_0 = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i^T \boldsymbol{w}) = \frac{1}{n}\sum_i y_i - \frac{1}{n}\sum_i x_i^T \boldsymbol{w} = \overline{y} - \overline{x}^T \boldsymbol{w} \Longrightarrow$ *Proved*

II 1.2 Solution:

Set new variables $l_1, l_2, ...l_n$:

min $\quad \frac{1}{n}\sum_{i=1}^{n} l_i$

s.t.

h(x$_i$) $- y_i \leq l_i$,

h(x$_i$) $- y_i \geq -l_i, i = 1, 2, ...n$

2. Convolution Layers (5 points)

We have a video sequence and we would like to design a 3D convolutional neural network to recognize events in the video. The frame size is 32x32 and each video has 30 frames. Let's consider the first convolutional layer.

- We use a set of $5 \times 5 \times 5$ convolutional kernels. Assume we have 64 kernels and apply stride 2 in spatial domain and 4 in temporal domain, what is the size of output feature map? Use proper padding if needed and clarify your notation.

- We want to keep the resolution of the feature map and decide to use the dilated convolution. Assume we have one kernel only with size $7 \times 7 \times 5$ and apply a dilated convolution of rate 3. What is the size of the output feature map? What are the downsampling and upsampling strides if you want to compute the same-sized feature map without using dilation?

Note: You need to write down the derivation of your results.

I 2.1 Solution: pad setting here is to get integer size of next layer

input data shape (video) = 32 x 32 x 32 x 3

kernel batch shape = 5 x 5 x 5 x 3 x 64

Because we know that for temporal domain stride s1 = 2, for domain stride s2 = 4

So according to 3D conv size function:

temp = (framesize-kernelsize+pad)/s2 + 1 = (30-5+3)/4 + 1 = 8

height = width = (inputsize - kernelsize + pad)/s2 + 1 = (32-5+1)/2 + 1 = 15

So the size of feature map is 8 x 15 x 15 x 64

II 2.2 Solution:

kernel size = 7 x 7 x 5

dilated convolution rate = 3

so after dilated convolution, the size 3 * (originsize-1) + 1 then size = 19 x 19 x 13

so the size of feature map should be the same as 2.1:

temp = (framesize-kernelsize+pad)/stride + 1 = (30-13+12)/1 + 1 = 30

height = width = (inputsize - kernelsize + pad)/stride + 1 = (32-19+18)/1 + 1 = 32

The size of feature map = 30 x 32 x 32 x 3 x 1

Downsampling = Upsampling = dilated rate = 3

3. Batch Normalization (5 points)

With Batch Normalization (BN), show that backpropagation through a layer is unaffected by the scale of its parameters.

- Show that

$$BN(\text{Wu}) = BN((a\text{W})\text{u})$$

where u is the input vector and W is the weight matrix, a is a scalar.

- (Bonus: 5 pts) Show that

$$\frac{\partial BN((a\text{W})\text{u})}{\partial \text{u}} = \frac{\partial BN(\text{Wu})}{\partial \text{u}}$$

I 3.1 Solution:

$BN(\text{Wu}) = \frac{\text{Wu} - E[\text{Wu}]}{\sqrt{Var[\text{Wu}]}} = \frac{a\text{Wu} - aE[\text{Wu}]}{aVar[\sqrt{\text{Wu}}]}$

$= \frac{(\text{aW})\text{u} - E[(\text{aW})\text{u}]}{Var[\sqrt{(\text{aW})\text{u}}]}$

$= BN((a\text{W})\text{u})$

II 3.2 Solution (Bonus):

According to equation in 3.1:

$\frac{\partial BN((a\text{W})\text{u})]}{\partial \text{u}} = \frac{\partial(\frac{a\text{Wu} - aE[\text{Wu}]}{aVar[\sqrt{\text{Wu}}]})}{\partial \text{u}}$

$= \frac{\partial(\frac{\text{Wu} - E[\text{Wu}]}{Var[\sqrt{\text{Wu}}]})}{\partial \text{u}}$

$= \frac{\partial BN(\text{Wu})}{\partial \text{u}} \implies Proved$