CS280 Fall 2018 Assignment 2
Part A

CNNs

Due in class, Nov 02, 2018

Name:

Student ID:

1. Linear Regression(10 points)

- Linear regression has the form $E[y|x] = w_0 + \boldsymbol{w}^{\boldsymbol{T}} x$. 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.

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.

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
$$BN(\mathbf{W}\mathbf{u}) = BN((a\mathbf{W})\mathbf{u})$$
where $\mathbf{u}$ is the input vector and $\mathbf{W}$ is the weight matrix, a is a scalar.

- (Bonus: 5 pts) Show that
$$\frac{\partial BN((a\mathbf{W})\mathbf{u})}{\partial \mathbf{u}} = \frac{\partial BN(\mathbf{W}\mathbf{u})}{\partial \mathbf{u}}$$