MATH 690: Topics in Data Analysis and Computation November 21, 2017

Convolutional neural network

Lecturer: Xiuyuan Cheng *Scriber: Zuoming Dai, Tianyu Wang*

1 Fully connected layers and convolutional layers

1.1 Fully connected layers

In a single layer of neural network, we (1) compute the linear combination of the input, (2) add a bias and pass the result to a non-linear activation function.

$$h_{\frac{1}{2}} = W\mathbf{x}, \quad \mathbf{x} \in \mathbb{R}^{d_x} \tag{1}$$

$$h = \sigma(h_{\frac{1}{2}}(\mathbf{x}) + \mathbf{b}). \quad h_{\frac{1}{2}}, \mathbf{b} \in \mathbb{R}^{d_h} \tag{2}$$

Now, rewrite $h_{\frac{1}{2}}$ into element wise summation,

$$h_{\frac{1}{2}}(u) = \sum_{u'=1}^{d_x} W(u', u)\mathbf{x}(u'), \quad u = 1 : d_h$$

In fully connected layers, $W(u', u)$ is a d_x-by-d_h dense matrix that connects each neurons in a layer to each neuron in the next layer. Each calculation of $h(u)$ costs $O(d_x)$ operations.

1.2 Convolutional layers

In a convolutional layers, W is supported by k places (a receptive field) where k is a constant. This makes $h_{\frac{1}{2}}$ a convolution.

$$W(u', u) = W(u' - u)$$
$$h_{\frac{1}{2}}(u) = \sum_{u'=1}^{d_x} W(u' - u)\mathbf{x}(u')$$

2 Models of convolutional neural network

2.1 1D signal + multiple channels

The 1D signal is a vector on each channel. Input data $\mathbf{x}(u, \lambda)$ has two parameters, where u is the spatial parameter and λ is the channel parameter, $u = 1 : N, \lambda = 1 : M$. Channels can separate the information of signals; for instance, a digital image that includes color information usually has 3 channels (red, green, blue). A visual illustration is given in Figure 1.

Fully-connected:

$$h_{\frac{1}{2}}(u, \lambda) = \sum_{\lambda', u'} W(u', u, \lambda', \lambda)\mathbf{x}(u', \lambda')$$

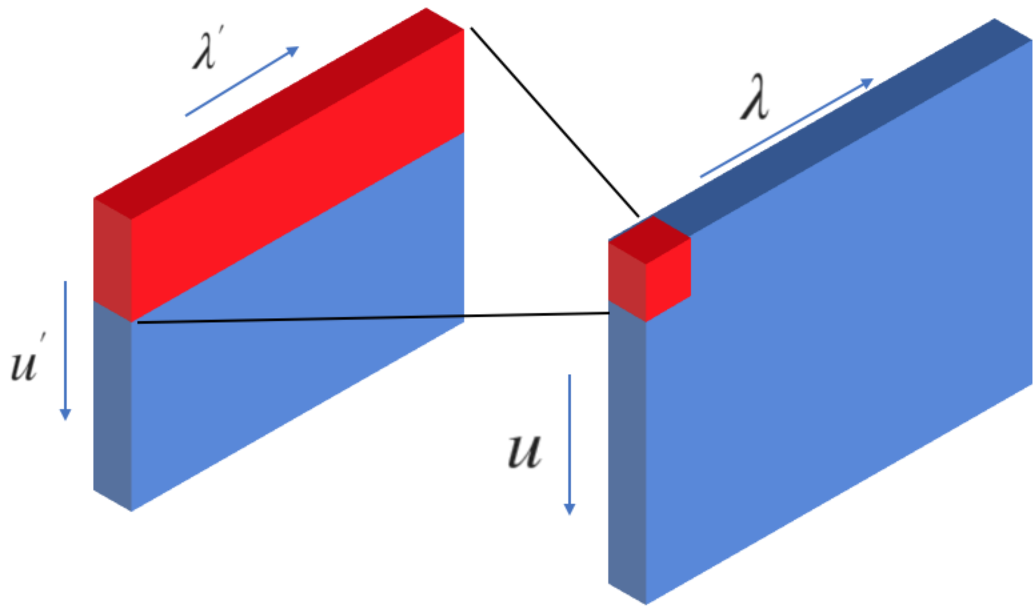

Figure 1:

Convolution in u:

$$h_{\frac{1}{2}}(u,\lambda) = \sum_{\lambda'} \sum_{u'} W(u'-u,\lambda',\lambda)\mathbf{x}(u',\lambda')$$
$$h_{\frac{1}{2}}(\cdot\,,\lambda) = \sum_{\lambda'} F_{\lambda',\lambda}(\cdot) * \mathbf{x}(\cdot\,,\lambda')$$
$$h(\cdot\,,\lambda) = \sigma(h_{\frac{1}{2}}(\cdot\,,\lambda) + \mathbf{b}(\lambda)),$$

where $F_{\lambda',\lambda}(\cdot) = W(\cdot\,,\lambda',\lambda)$ is the spatial operator.

Remark 1. *The bias* $\mathbf{b}(\lambda)$ *is always constant among space.*

2.2 2D signal + multiple channels

A 2D signal $\mathbf{x}(\vec{u},\lambda)$ has 2 dimensions for spatial information, $\vec{u} \in [N] \times [N]$, and one dimension for channel information, $\lambda = 1 : M$. In the continuous version, $\vec{u} \in [0,1]^2$ or $\vec{u} \in \mathbb{R}^2$.

$$h(\vec{u},\lambda) = \sum_{\lambda'} \sum_{\vec{u}'} W_{\lambda',\lambda}(\vec{u}'-\vec{u})\mathbf{x}(\vec{u}',\lambda')$$

where $W_{\lambda',\lambda}(\cdot)$ is the filter supported on a patch window. A visual illustration is given in Figure 2.

Remark 2. *Count of parameters of weights in one layer:*

In a fully-connected neural network, $W = W(\vec{u}',\vec{u},\lambda',\lambda)$*, where* $\vec{u}',\vec{u},\lambda',\lambda$ *consume the space of* N^2, N^2, M', M *respectively. Therefore, the total number of the parameter stored is* $N^2 \cdot N^2 \cdot M' \cdot M$.

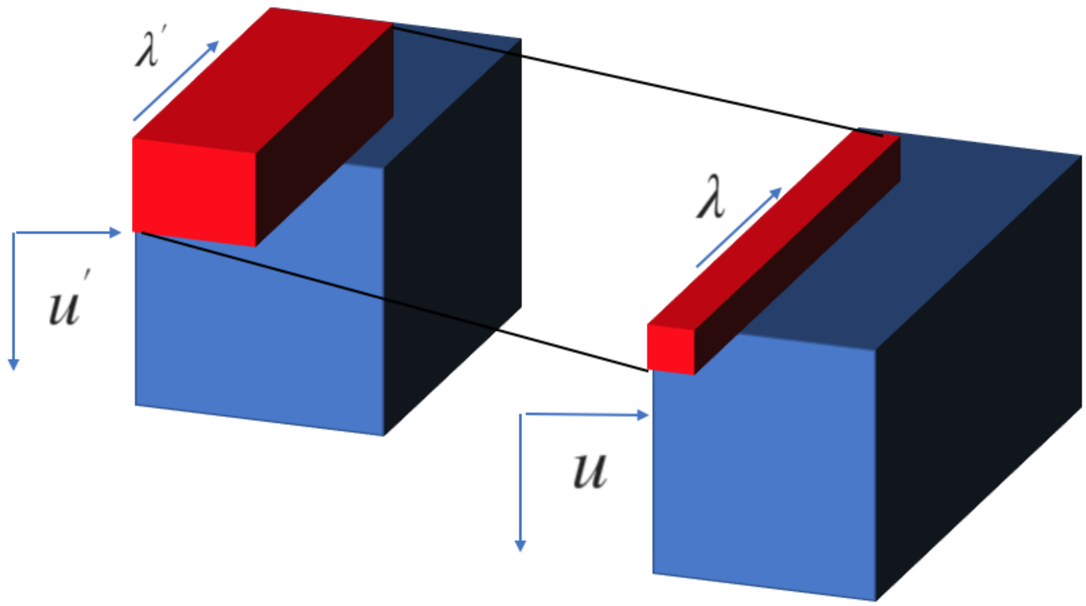

Figure 2:

In a convolutional neural network, suppose $W = W(\vec{u}' - \vec{u}, \lambda', \lambda)$ is supported on a k-by-k patch, where $\vec{u}' - \vec{u}, \lambda', \lambda$ consume the space of k^2, M', M respectively. Therefore, the total number of the parameter stored is $k^2 \cdot M' \cdot M$.

3 Instability and stability of CNN

The goal of stability is to stabilize the output f_θ of the convolutional neural network against a small natural perturbation to the input data. In other words, we hope that $D(f_\theta(\mathbf{x}), f_\theta(\mathbf{x}'))$ is small if $d(\mathbf{x}, \mathbf{x}')$ is small, where D, d are the distances in output space and input space respectively.

3.1 L^2 stability

Suppose $f_\theta : \mathbb{R}^{d_x} \to \mathbb{R}^{d_y}$ is the model of a convolutional neural network. Model f_θ has L^2 stability, if and only if there exists a positive constant c such that for all input $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^{d_x}$,

$$||f_\theta(\mathbf{x}) - f_\theta(\mathbf{x}')||_2 \leq c \cdot ||\mathbf{x} - \mathbf{x}'||_2.$$

Remark 3. *There are many ways to achieve L^2 stability, e.g. controling the magnitude of $|W_{ij}|, ||\nabla_{\mathbf{x}} f_\theta(\mathbf{x})||$*

For example, it is easy to observe that Proposition 1 holds.

Proposition 1. *If $|W_{ij}|$ is bounded, then f_θ has L^2 stability.*