Homework Advanced Learning Models 2018

MSIAM - MOSIG

Julien Mairal & Jakob Verbeek

Guidelines

- Homeworks are due by January 7, 2019.
- Send your worked-out solutions in a latex-generated PDF to `dexiong@inria.fr`
- You can work in a group of two people, but for the data challenge you cannot work with the same person again.
- The homework makes out 30% of your final grade.

1 Convolutional Neural Networks

1.1 Parameter count

Suppose we have a convolutional layer in a network that maps a feature map of size 32×32 pixels and 24 feature channels to an output of size 32×32 pixels with 32 feature channels, using a filter of size 5×5. What is the number of parameters of this layer ?

1.2 Computational cost

Suppose we have a convolutional layer in a network that maps a feature map of size $W \times H$ pixels and N feature channels to an output of size $W \times H$ pixels with M feature channels, using a filter of size $w \times h$. What is the computational cost (in terms of additions and multiplications) of this layer in "big-O" notation?

1.3 Feature map sizes

Suppose we have a convolutional network with four layers, with filters of size 3×3, 7×7, 5×5, and 9×9 from first to last layer. Suppose there is no padding used in the network, and that the input is 87×132 pixels. What is the size of the network output?

1.4 Feature map sizes

What if the second and fourth layer use a stride of 2 pixels?

2 Multi-layer perceptron

Suppose we have a small MLP with one hidden layer of 3 neurons, 2 input variables, and one output unit. The hidden layer uses a ReLU activation.

The weights are as follows:

- From input to hidden unit 1: $(0, -1)$
- From input to hidden unit 2: $(1, 1)$
- From input to hidden unit 3: $(-1, 1)$
- From hidden to output $(-2, -1.5, -1.5)$.

There are no biases in the network. The output layer does not have a non-linearity.

2.1 Forward propagation

Compute the output of the network for the following five inputs: (0,0), (4,0), (-4,0), (0,4), (0,-4).

2.2 Back propagation

Suppose we use the squared loss $l(y,t) = \frac{1}{2}(y-t)^2$ between the scalar output of the network y and the target value t. Let the targets of the inputs above be 0, -10, -10, -10, -10.

For each of the five inputs, compute the gradients of the loss with respect to the two weight matrices between input and hidden, and hidden and output. Assume the derivative of the ReLU at zero equals zero.

3 Kernel examples

Are the following kernels positive definite? Provide a proof in each case

1.

$$\forall x, y \in \mathbb{R} \quad K_1(x,y) = 10^{xy}\,, \quad K_2(x,y) = 10^{x+y}\,.$$

2.

$$\forall x, y \in [0,1) \quad K_3(x,y) = -\log(1-xy)\,.$$

3. Let $\mathcal{X}$ be a set and $f, g : \mathcal{X} \to \mathbb{R}_+$ two non-negative functions:

$$\forall x, y \in \mathcal{X} \quad K_4(x,y) = \min(f(x)g(y), f(y)g(x))$$

4 Combining kernels

1. For $x, y \in \mathbb{R}$, let

$$K_1(x,y) = (xy+1)^2 \quad \text{and} \quad K_2(x,y) = (xy-1)^2\,.$$

What is the RKHS of K_1? Of K_2? Of $K_1 + K_2$?

2. Let K_1 and K_2 be two positive definite kernels on a set $\mathcal{X}$, and α, β two positive scalars. Show that $\alpha K_1 + \beta K_2$ is positive definite, and describe its RKHS.