Assignment 2 (Due: Nov. 17, 2019)

1. (Math) In our lecture, we mentioned that for logistic regression, the cost function is,

\[
J(\theta) = -\sum_{i=1}^{m} y_i \log(h_\theta(x_i)) + (1 - y_i) \log(1 - h_\theta(x_i))
\]

Please verify that the gradient of this cost function is

\[
\nabla_\theta J(\theta) = \sum_{i=1}^{m} x_i (h_\theta(x_i) - y_i)
\]

2. (Design) In our lecture, we mentioned NIN architecture which introduces a “mlpconv” concept. The mlpconv layer maps the input local patch to the output feature vector with a multilayer perceptron (MLP) consisting of multiple fully connected layers with nonlinear activation functions. The following figure compares the linear convolution layer with the MLPconv layer. For more details, you can refer to the paper “M. Liu et al., Network in network, in Proc. ICLR, 2014”.

![Linear convolution layer vs MLPconv layer](image)

Actually, mlpconv can be implemented with conventional convolution layers. So, please specify the network design to implement the following particular mlpconv layer.

For a mlpconv layer, suppose that the input feature map is of the size $m \times m \times 32$, the expected output feature map is of the size $m \times m \times 64$, the receptive field is $5 \times 5$; the mlpconv layer has 2 hidden layers, whose node numbers are 16 and 32, respectively.

**How to implement this mlpconv layer with convolutional layers?**
3. (Math) In the lecture, we talked about the least square method to solve an over-determined linear system. $Ax = b, A \in \mathbb{R}^{m \times n}, m > n, rank(A) = n$, the closed form solution is $x = (A^T A)^{-1} A^T b$. Try to prove that $A^T A$ is non-singular (or in other words, it is invertible).