Homework Assignment 3

Kexin Huang, kh2383

March 9, 2018

1. Cross-validation is a useful strategy for model selection, especially when the training data is small. However, it cannot be used for early-stopping (in other words, you cannot pick the best fold). Why is this the case?

Because each fold has different validation dataset, early stopping pick the classifier with least validation cost computed from a single same validation dataset. Therefore, it is not comparable among each fold as the least validation cost classifier pick on one fold has a different validation cost on another fold.

2. In multi-class classification, given the definitions in the lecture notes, derive the following distance function. defined as

$$\begin{aligned} D(y^*, M, \mathbf{x}) &= -\log p_{M^*(\mathbf{x})} \\ &= -a_{y^*} + \log \sum_{k=1}^{K} \exp(a_k), \end{aligned}$$

In multi-class classification, the distance function is defined to be the negative log likelihood of the correct category. Suppose the correct category has index x. And we denote C as the random variable from event space to a measurable space E which contains all the indices. There are K categories. And denote $a_i = W_i X$ the i th category's score. So the probability P (C=x) is defined as $\frac{exp(a_x)}{\sum_{k=1}^{K} exp(a_k)}$ Therefore, the distance function is

D(y*,M,x) = -log(P(C=x)

= -log($\frac{exp(a_x)}{\sum_{k=1}^{K} exp(a_k)}$)

= -[$log(exp(a_x)) - \log(\sum_{k=1}^{K} exp(a_k))$]

=$-[a_x - \log(\sum_{k=1}^{K} exp(a_k)]$

=$-a_x + \log(\sum_{k=1}^{K} exp(a_k)$

and because y* is the correct category index, so $x = y^*$

So, D(y*,M,x) = $-a_{y*} + \log \sum_{k=1}^{K} exp(a_k)$

3. Given the definition of the distance function above, derive a learning rule step-by-step for each column vector $\mathbf{w}_c$ of the weight matrix $\mathbf{W}$ (Equation 1.28 in the lecture notes).

We know from last task the distance function is D(y*,M,x) = $-a_{y*} + \log \sum_{k=1}^{K} exp(a_k)$, we want to derive the partial derivative of row vector Wy respect to D. that is $\frac{\partial D}{\partial W_y}$

Let's denote $-a_y$ as D1 and $\log \sum_{k=1}^{K} exp(a_k)$ as D2. So $\frac{\partial D}{\partial W_y} = \frac{\partial D1}{\partial W_y} + \frac{\partial D2}{\partial W_y}$

$\frac{\partial D2}{\partial W_y} = \frac{\partial \log \sum_{k=1}^{K} exp(a_k)}{\partial W_y} = \frac{\partial \log \sum_{k=1}^{K} exp(a_k)}{\partial \sum_{k=1}^{K} exp(a_k)} * \frac{\partial \sum_{k=1}^{K} exp(a_k)}{\partial W_y} = \frac{1}{\sum_{k=1}^{K} exp(a_k)} * \sum_{k=1}^{K} [\frac{\partial exp(a_k)}{\partial W_y}] =$
$\frac{1}{\sum_{k=1}^{K} exp(a_k)} * \sum_{k=1}^{K} [\frac{\partial exp(a_k)}{\partial a_k} * \frac{\partial a_k}{\partial W_y}] = \frac{1}{\sum_{k=1}^{K} exp(a_k)} * \sum_{k=1}^{K} [exp(a_k) * \frac{\partial a_k}{\partial W_y}]$

$a_y = W_y * X$, therefore, $\frac{da_y}{dW_y} = \frac{dW_y * X}{dW_y} = X$ and $\frac{da_k}{dW_y} = \frac{dW_k * X}{dW_y} = 0$ when y not equal to k. Therefore, $\sum_{k=1}^{K} [exp(a_k) * \frac{\partial a_k}{\partial W_y} = 0 + 0 + .. + exp(a_y) * X + 0 + .. + 0 = exp(a_y) * X$, therefore, the original equation becomes $\frac{1}{\sum_{k=1}^{K} exp(a_k)} * exp(a_y) * X$ and we know $\frac{1}{\sum_{k=1}^{K} exp(a_k)} * exp(a_y) = p(c = y|X)$, therefore, the original equation $\frac{\partial D2}{\partial W_y}$ becomes $p(c = y|X) * X$

And we already proved that $\frac{\partial -a_k}{\partial W_y} = -X$ when k=y, and =0 when k not equal to y. Therefore, when y corresponding to the correct class, k=y, the $\frac{\partial D}{\partial W_y} = \frac{\partial D1}{\partial W_y} + \frac{\partial D2}{\partial W_y} = -X + p(c = y|X) * X = -(1 - p(c = y|X)) * X$
and when y corresponding to a wrong class, k!=y, $\frac{\partial D}{\partial W_y} = \frac{\partial D1}{\partial W_y} + \frac{\partial D2}{\partial W_y} = -0 + p(c = y|X) * X = -(0 - p(c = y|X)) * X$

end of proof

4. Multiclass Classification on MNIST Please download `https://github.com/nyu-dl/Intro_to_ML_Lecture_Note/blob/master/homeworks/hw3.ipynb` and follow its instructions.

please see attached code file