進階機器學習 Advanced Machine Learning

Homework #1
Due 2023 Mar 8 11:00PM

(一) For the 20 pairs of (x,y) data provided in the attached ***HW1-1.csv*** file, use either tensorflow2.0 or pytorch to implement：

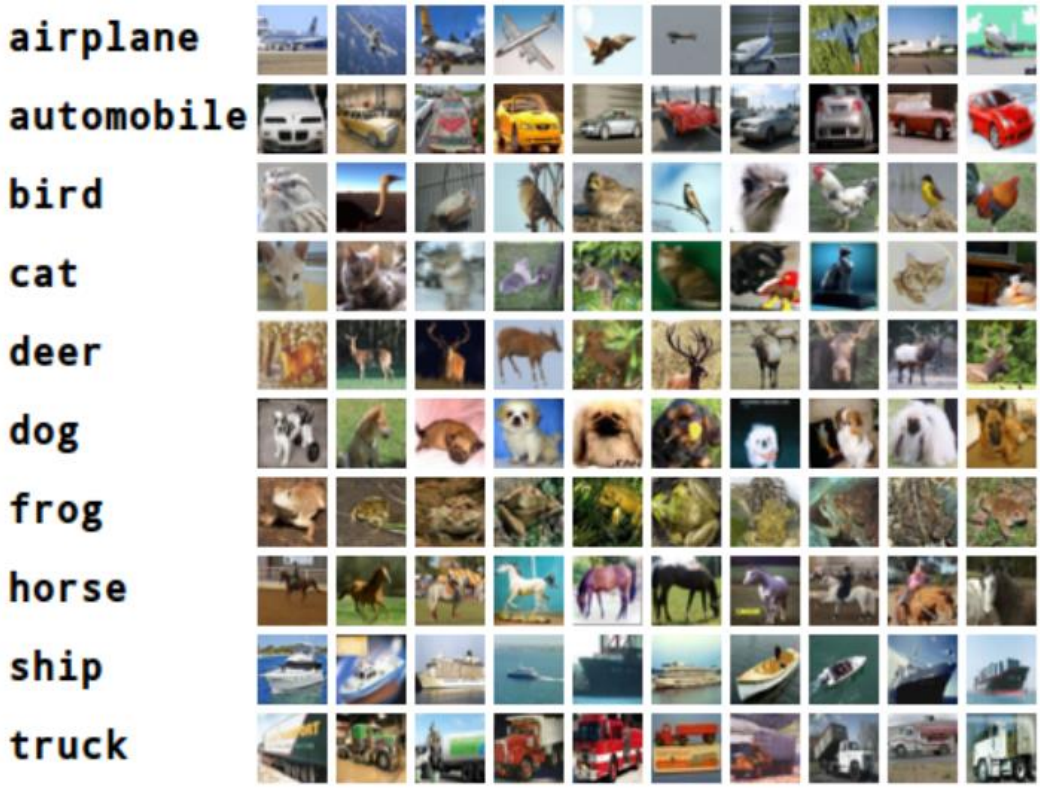
(a) Writing a multi-layer neural network model to simulate a function y=f(x) which can produce the above data. The output data y produced by the neural network for a given input data x must have a mean error (MSE) less than 0.1.

(b) Write a multi-order regression equation model $y = w_m x^m + \cdots + w_1 x^1 + w_0$ to also simulate the above function with the same MSE requirement. You should

(i) Print out the coefficients of the equation $w_m, \ldots, w_0$。

(ii) Use the tensorflow or pytorch to calculate the following gradients: $f'(3.0), f'(0.1), f'(-0.5)$.

(二) Cifar-10 dataset consists of 60,000 RGB color pictures of 32*32, a total of 10 categories. It contains a total of 50,000 training data, and a total of 10,000 testing data.



Follow the following steps below to build a CNN model.

type	Kernel size(or pooling size)	Output channel	stride	Padding
Conv2D	3x3	32	1x1	The same
relu				

MaxPooling	2x2	--	2x2	The same
Dropout	25%			
Conv2D	3x3	64	1x2	The same
Relu				
MaxPooling	2x2	--	2x2	The same
Dropout	25%			
Flatten				
Dense(fully connected)	Output size=1024			
Dropout	25%			
Dense(fully connective)	Output size=1024			
Dropout	25%			
Dense(fully connected)	Output size=10			

Answer the following questions:

(a) Calculate the total number of parameters (including kernel and bias). List the calculation process. You compare it with the results of ***model.summary()*** in tensorflow or ***torchsummary(model)*** in pytorch.

(b) Calculate the top-1 and top-5 accuracy and loss of the training & validation data.

(c) Adjust the architecture by yourself (increase/decrease the number of convolution, max pooling, relu, dropout, and internal parameters), so that the top-1 accuracy result of the test data is more than 10% higher than the above architecture.

(三)Suppose we have an input feature map with a size of 4x4, and the features will pass through a convolution layer with a kernel size of 3x3. The activation function used in convolution is *sigmoid*. The convolution results will be flattened and multiplied by the corresponding weights. The products will be added together to produce a final regression value. Assume the correct label value is yy, and the MSE has been used as the loss function. Derive the loss gradient with respect to for the weight *w10* in the following figure.