Deep Learning for Big Visual Data, Summer 2024, BUPT

Homework Assignment 1

Due 5:00 pm, Thursday, July 4th, 2024

1. (20%) Based on the lectures in the classes and your personal opinions, please list 3 most important reasons WHY deep learning based machine learning methods can often achieve better performance than the traditional machine learning methods?

2. (20%) Given a trained classifier for 4 object classes (C_1, C_2, C_3 C_4), an input data belongs to C3 generates (0.2, 0.1, 0.6, 0.1) output likelihood, what are the corresponding loss values (L1 loss, L2 loss MSE and cross-entropy loss.) associated with this data?

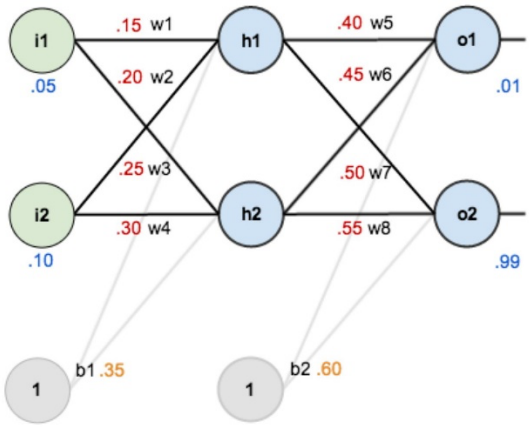
3. (20%) Given the **confusion matrix** of this 4-class classifier

		Predicted Classes			
Actual Classes		C_1	C_2	C_3	C_4
	C_1	68	12	9	11
	C_2	14	74	5	7
	C_3	12	3	82	3
	C_4	6	10	12	72

Please compute the (1) overall average accuracy. Followed by (2) per class precision and recall, (3) F1 score of C4, and the (4) micro-average precision of all 4 classes.

4. (20%) Based on the procedures and the values of computing the MSE minimized backpropagation updated weights w5, please compute the updated value of w2 of the following 2-layer MLP with given initialized weights (in red) and the bias offsets (in orange), where the 2-D inputs are (0.05, 0.10), and the desired/target outputs are (0.01, 0.99). All neurons are assumed to have sigmoid nonlinear activations, and the learning rate is 0.5.

   Note that you need to compute both $\frac{\partial Etotal}{\partial outo1}$ and $\frac{\partial Etotal}{\partial outo2}$ to compute $\frac{\partial Etotal}{\partial outh1}$ before you can compute $\frac{\partial Etotal}{\partial w2}$



Sigmoid 函数式: $$g(z) = \frac{1}{1+e^{-z}}$$

5. (20%) Given one Section of Google Inception CNN architecture shown below, please compute the total number of connections and the trainable convolution parameters? In this Section, the input feature maps (previous layer) are of size 28x28, with channel sizes of 192. The output concatenated feature maps are also of size 28x28, with 256 channels. Various convolution kernels are used as shown in the Figure, please compute all the trainable parameters used in this Section?