Adaptive Computation and Machine Learning
COMS 4030A   and   COMS 7047A

TEST   10 April 2025

Time: 90 minutes

Question 1. (*7 marks*)

Consider a neural network that is being trained using sum-of-squares loss function

$$L = \tfrac{1}{2} \sum_{\ell} (y_\ell - t_\ell)^2,$$

where ℓ ranges over all output nodes.

Recall that for a node n, we set $\delta_n = \frac{\partial L}{\partial z_n}\big|_{\boldsymbol{xtW}}$ (using the notation from the Lecture Notes).

(a) (*3 marks*) If n is an output node of the neural network, show that

$$\delta_n = (a_n - t_n) \left(\tfrac{dy_n}{dz_n}\big|_{\boldsymbol{xtW}} \right).$$

(b) (*4 marks*) If n is a node in a hidden layer in the neural network, show that

$$\delta_n = \left(\sum_{\ell} \delta_\ell \, \underline{w}_{n\ell} \right) \left(\tfrac{dy_n}{dz_n}\big|_{\boldsymbol{xtW}} \right)$$

where the ℓ in the summation ranges over all nodes ℓ in the layer after n's.

Question 2. (*3 marks*)

Explain what the cross-entropy loss function is and how it is used in training neural networks.

Question 3. (*3 marks*)

Explain what batch gradient descent is and describe how it works.

Question 4. (5 *marks*)

Use the entropy method to create a decision tree with the dataset S given below.
Show all your calculations and draw the final decision tree.

S :

F_1	F_2	target
t	x	Y
f	y	N
t	z	Y
t	x	N
f	z	Y
f	x	N
t	y	N
t	x	Y
f	z	Y
f	y	N
f	y	Y

Question 5. (3 *marks*)

Compute $\text{Gain}(T, F_1)$ for the dataset T below using the mean square error (MSE) and with split point of 0.75 for attribute F_1. Show all your calculations.

T :

F_1	target
0.3	1.2
0.7	1.6
0.5	0.9
0.1	0.7
0.9	1.3
0.8	0.9

Question 6. (4 *marks*)

(a) (3 *marks*) Describe the process of creating a random forest for a given dataset.
State the hyperparameters that are required for the process.

(b) (1 *mark*) Describe how a random forest is used to predict the target value of a given input.

Total marks: 25