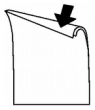
CSCE 5063 Fall 2018 Midterm 2

1. (9 points) I plan to return these tests face-down, so please write your name on the back, near the top.



2. (15 points) Given a univariate linear model, $h(x) = mx + b$, and the objective function $e = (y_i - h(x_i))^2$, if you present the training pattern $x_1 \rightarrow y_1$ to update this model by stochastic gradient descent with a learning rate of η (and no momentum), what will be the updated values for the parameters m and b? (Please try to make it easy for me to follow your work. Maybe consider first trying it on the back of another page, then copying the relevant parts to here.)

3. (15 points) Given a univariate linear model, $h(x) = f(mx + b)$, and the objective function $e = |y_i - h(x_i)|$, if you present the training pattern $x_1 \rightarrow y_1$ to update this model by stochastic gradient descent with a learning rate of η (and no momentum), what will be the updated values for the parameters m and b? (Note that the derivative of $f(x)$ can be expressed as $f'(x)$.)

4. (9 points) Please briefly identify 3 ways to limit over-fit with a neural network:

1.

2.

3.

5. (10 points) Given a one-dimensional input image with 4 elements, $\langle x_1, x_2, x_3, x_4 \rangle$, and a filter with 3 elements, $\langle w_1, w_2, w_3 \rangle$, please give two expressions for the two output values that would be obtained by convolving $X \oplus W$. (You may assume there is no bias term.)

6. (2 points) Draw a neural network diagram (with 4 input units and 2 output units) to represents the operation in the previous problem. Use a dotted line for w_1, a straight line for w_2, and a squiggly line for w_3.

7. (5 points) Assuming the blame on the output units is $\langle b_1, b_2 \rangle$, give an expression for the blame that should be assigned to the input (or the output of the previous layer).

8. (5 points) Give three expressions for updating the three weights in the filter using a learning rate of η.

9. (30 points) Please match these descriptions with their terms. (Write the letter in the blank space. Each term is used only once. I filled in a few of them for you.)

___ Helps with the problem of overfit.

___ Helps with the problem of noisy training data by smoothing the gradient.
___ Helps with the problem of vanishing/exploding gradients.
_A_ Helps with the problem of local optima.

_Z_ Can learn when all hypotheses are equally likely.
___ Evaluates a learning algorithm.

___ Evaluates a hypothesis (current set of weights).

___ Evaluates an input vector or feature vector.

___ Seeks the inputs that lead to desired outputs.
_B_ Trains an Elman network.

_M_ A model that takes labels as inputs and predicts features.
_E_ A model that takes features as inputs and predicts labels.
___ The main body of a neuron.

___ The connection that joins an axon to a dendrite. Analogous with weights.
___ A biologically-inspired learning model.

_W_ A structure of low dimensionality embedded in a higher-dimensional space.
_P_ Values that converge as an algorithm runs.

___ What gradient descent descends.

___ A common property in activation functions.

_L_ Regression with a model that has one linear layer and a sigmoid-shaped activation function.
___ Fits the training data well, but generalizes poorly.
___ More training samples than parameters in the model.
_D_ A map of all the neural connections in a nervous system.
___S-shaped activation function

A. Random restarts

B. Backpropagation through time

C. Momentum

D. Connectome

E. Generative model

F. Model

G. Objective function

H. Cross-validation

J. Soma

K. Synapse

L. Logistic regression

M. Discriminative model

N. Overfit

P. Latent

Q. Overconstrained

R. Careful weight initialization

S. Spiking neural network

T. Inference

U. Sigmoid

V. Regularization

W. Manifold

X. Error surface

Y. Monotonic

Z. Nothing