MA 225 MULTIVARIABLE CALCULUS
APPLICATION OF GRADIENT VECTOR TO AI

SIU-CHEONG LAU

Deep learning is based on a model of the neural network in human brain. Recall from the last optional assignment that a neuron is modeled by the function

$$z = \sigma(\vec{w} \cdot \vec{x} - b).$$

The vector $\vec{x}$ is composed of, say, two input numbers. (In practice it is composed of many inputs, for instance the number of pixels in a photo.) The vector $\vec{w} = (w_1, w_2)$ is a given weight vector associated to the neuron. b is a given threshold value. (Think of them as controlled by chemicals inside the neuron.)

$$\sigma(y) = \frac{1}{1 + e^{-10y}}$$

is an approximation of a step function. Then z is the output of the neuron according to the above formula. (In practice we should consider layers of neurons rather than just one neuron.)

The neural network is trained by a given set of inputs and 'correct answers'. (For instance the inputs can be human voice, and the correct answers can be whether 'yes' or 'no'.) Then the neural network learns by adjusting the weight vectors $\vec{w}$ and threshold values b, so that its output vector $\vec{z}$ gets closer to the correct answer $\vec{z}_{\text{correct}}$.

The squared distance (magnitude of error) is

$$E = \|\vec{z} - \vec{z}_{\text{correct}}\|^2.$$

Suppose the given input and answer are $\vec{x} = (1, 0)$ and $\vec{z}_{\text{correct}} = 1$ respectively. Let's say the current weight and threshold are $\vec{w} = (0.2, 0.8)$ and $b = 0.1$.

Question A: We regard E as a function in the variables (w_1, w_2, b). Find the partial derivatives

$$\frac{\partial E}{\partial w_1}, \frac{\partial E}{\partial w_2}, \frac{\partial E}{\partial b}$$

at $(w_1, w_2, b) = (0.2, 0.8, 0.1)$?

Question B: What is the gradient vector $\text{grad} E|_{(w_1,w_2,b)=(0.2,0.8,0.1)}$?

Question C: How should the neuron change its weight vector $\vec{w}$ and the threshold b? (Recall that the gradient vector field is pointing to the direction that E increases most rapidly. We want to decrease the error E.)

Thus the key part of the deep learning algorithm is to compute gradient vectors!