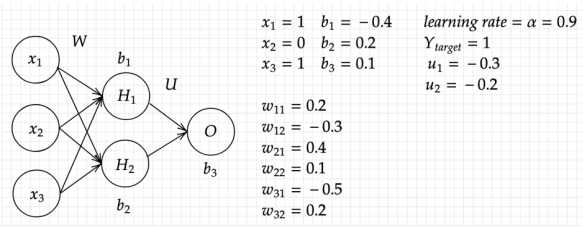
Machine learning
Fall 2023
HW2

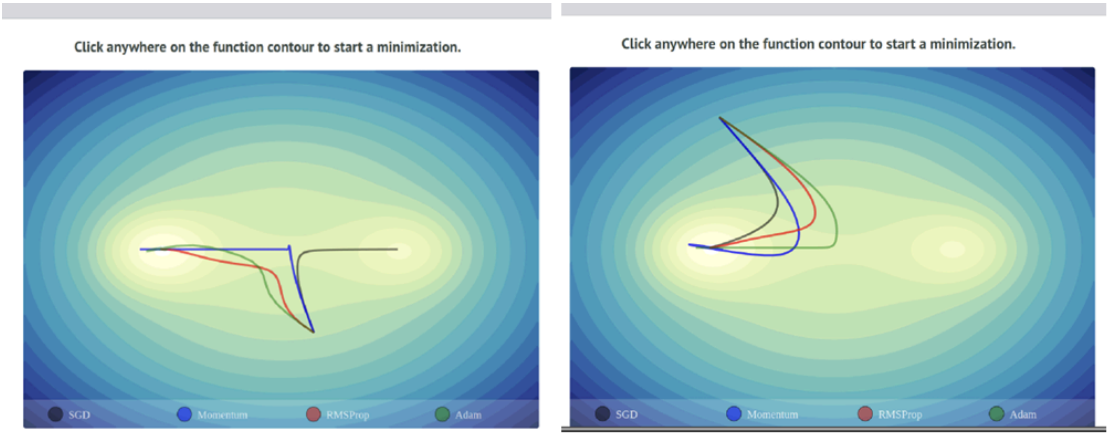
---

1. In the neural network below, which has one hidden layer, perform the backpropagation algorithm twice and update the weight vectors in the network using the gradient descent method along with momentum. The initial values are shown in the figure. Use the mean squared error function as the cost function. The activation functions are sigmoid functions.

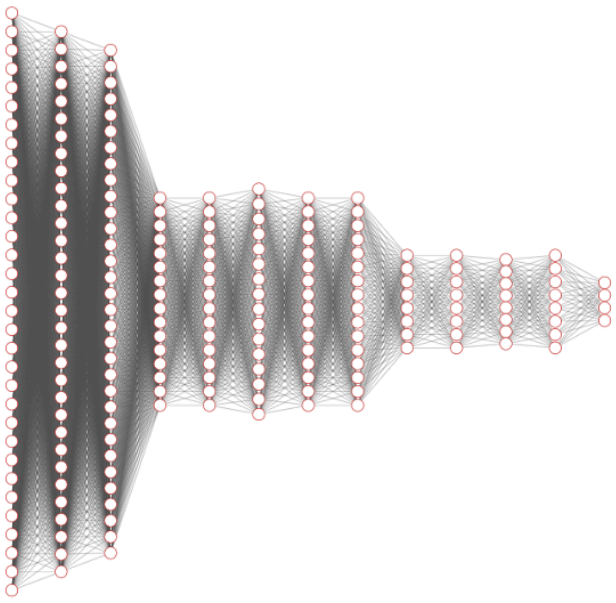


2. Please refer to this link, which is a visualization tool for optimization algorithms discussed in class, and choose two starting points for running the algorithms. At least one of the endpoints of the algorithm runs in these two scenarios should be different. Provide the necessary mathematical equations for these algorithms, along with the final points and their speeds compared to each other, and analyze and compare them based on these equations.

for instance we have these two senario's:



3. We have a neural network as shown below, which is intended to predict a multi-class problem for us. We trained it using the materials we have learned so far in the course. However, to our surprise, we found that the weight vectors obtained after training the network for 1000 iterations are unable to predict with high accuracy, and the derivative vectors of most of these weights are close to zero. What do you think is the main reason for this issue? Explain your solution.



4. Define a nonlinear logistic regression model (based on the decision boundary function $u(x)$) with on the following information:

- The training data $D = \{ (x_k , c_k) \mid k = 1, 2, ..., N \}$, where $c_k \in \{0, 1\}$ and x_k is a numerical value.

- The decision boundary function is defined as:

$$u(x) = w_2 x^2 + w_1 x + w_0$$

Write the total cost function and the gradient of the cost function with respect to the weight vector $w = \begin{bmatrix} w_2 & w_1 & w_0 \end{bmatrix}$.