In the name of GOD.

Machine Learning

Spring 2025

Sharif University of Technology
Hamid R. Rabiee, Zahra Dehghanian

Homework 6 | Deadline: 1404/03/20

1. Consider the following neural network. Single-circled nodes denote variables (e.g., x_1 is an input variable, h_1 is an intermediate variable, $\hat{y}$ is an output variable), and double-circled nodes denote functions (e.g., Σ takes the sum of its inputs, and σ denotes the logistic function $\sigma(x) = \frac{1}{1+e^{-x}}$). In the network below,

$$h_1 = \frac{1}{1 + e^{-x_1 w_1 - x_2 w_2}}.$$

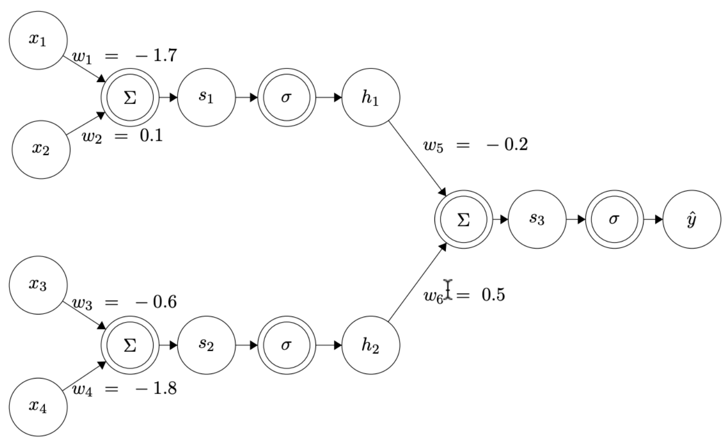


Figure 1: Neural Net architecture

Suppose we use an L2 loss function:

$$L(y, \hat{y}) = \|y - \hat{y}\|_2^2.$$

We are given a data point $(x_1, x_2, x_3, x_4) = (-0.7,\ 1.2,\ 1.1,\ -2)$ with true label $y = 0.5$. Use the backpropagation algorithm to compute the partial derivative: $\frac{\partial L}{\partial w_1}$.

2. The standard cost function for k-means clustering is defined as:

$$L = \sum_{j=1}^{k} \sum_{x_i \in S_j} ||x_i - \mu_j||_2^2$$

where S_j is the set of data points x_i assigned to cluster j, and μ_j is the centroid of cluster j. Answer the following:

(a) **Optimality of the Centroid and Distance Metrics:**

i. Prove that for a given set of points S_j assigned to a cluster j, the unique vector μ_j that minimizes $\sum_{x_i \in S_j} ||x_i - \mu_j||_2^2$ is the sample mean of the points in S_j.

ii. How would the optimal cluster center μ_j change if the L_2 norm (Euclidean distance squared) in the cost function were replaced with the L_1 norm (Manhattan distance), i.e., $L' = \sum_{j=1}^{k} \sum_{x_i \in S_j} ||x_i - \mu_j||_1$? Provide a characterization of this new optimal center (e.g., geometric median).

(b) **Convergence and Initialization:**

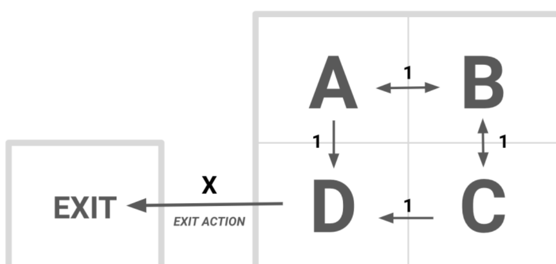
i. The k-means objective function L is non-convex. Explain the implications of this non-convexity for the convergence of the algorithm. Why does the standard k-means algorithm (Lloyd's algorithm) guarantee convergence to a local minimum but not necessarily the global minimum?

(c) **Relationship to Gaussian Mixture Models (GMMs):**

i. What specific assumptions must be made about the covariance matrices and mixing coefficients of a GMM for its EM algorithm to effectively reduce to the k-means algorithm?

ii. Explain how the "hard" assignment of k-means (each point belongs to exactly one cluster) contrasts with the "soft" assignment in GMMs.

3. Consider below Markov Decision Process (MDP) with four states: A, B, C, and D. From states A, B, and C, the agent can choose from the actions: `LEFT`, `RIGHT`, `UP`, or `DOWN`, unless obstructed by a wall in that direction. From state D, the only available action is a special `EXIT` action, which grants the agent a terminal reward of x. All other actions (non-exit actions) yield a reward of 1.



(a) **(Deterministic actions)**
Assume all actions are deterministic, and the discount factor is $\gamma = \frac{1}{2}$. Express the value function $V^*(s)$ for the following states in terms of x, given that the optimal value function satisfies:

$V^*(D) =$ $\qquad$ $V^*(A) =$

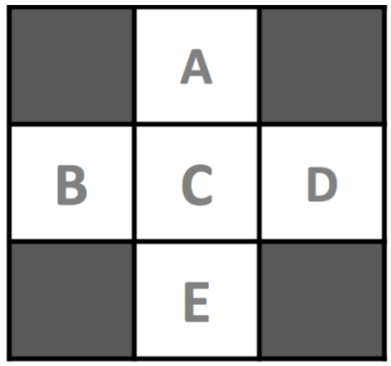
$V^*(C) =$ $\qquad$ $V^*(B) =$

(b) **(Stochastic actions)**
Now suppose each non-exit action succeeds with probability $\frac{1}{2}$; otherwise, the agent remains in the same state and receives a reward of 0. The `EXIT` action from state D is still deterministic and always succeeds. Let $\gamma = \frac{1}{2}$ as before.

Find the value of x for which the agent is indifferent between two actions from state A: taking action `DOWN` to go to D, and taking action `RIGHT` to go to B. In other words, solve for x such that:

$$Q^*(A, \texttt{DOWN}) = Q^*(A, \texttt{RIGHT})$$

4. Consider a Gridworld scenario where an agent aims to estimate the value function of each state using TD Learning and Q-Learning.



Suppose we observe the following $(s, a, s', R(s, a, s'))$ transitions and rewards (in order from left to right):

(B, East, C, 2), (C, South, E, 4), (C, East, A, 6), (B, East, C, 2)

Assume the initial value of each state is 0, the discount factor $\gamma = 1$, and the learning rate $\alpha = 0.5$.

(a) What are the learned state values from TD learning after all four transitions?

(b) What are the learned Q-values from Q-learning after all four observations? Use the same $\alpha = 0.5$ and $\gamma = 1$.