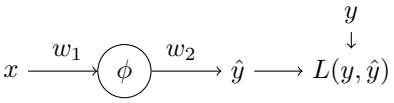
1 Question



Consider the simple neural network in the figure above. The non-linear activation ϕ ist the ReLU activation, given as $\phi(z) := \max(\ ,0)$ and the loss function L is squared error: $L(y, y) : (y - \hat{y})^2$.The weights are given as $w1 = 1$ and $w2 = 3$. Note that no biases are used.

1.1

Compute the loss L for input $a = 2$ and target $y = 5$. Provide all intermediate computations. [3points]

2 Question

Consider a linear regression problem with data $(x_1, y_1) = (1, 3)$, $(x_2, y_2) = (2, 5)$ and $(x_3, y_3) = (3, 4)$.

2.1

Construct the design matrix X, containing a constant "dummy feature" of 1 (for the bias) and the input features. Also construct the vector y of regression targets.

2.2

State the formula for the analytic least-squares solution and use it to compute optimal parameters θ^*.Furthermore, compute the least-squares loss for θ^*. Provide intermediate computations Hint: The inverse of a 2×2 matrix is given as $\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$

2.3

Sketch the data in a coordinate system and draw the graph of the learned regression function.

3

Consider the following data set:

$$\boldsymbol{x^{(1)}} = (1, 2)^T \quad \boldsymbol{x^{(2)}} = (3, 2)^T \quad \boldsymbol{x^{(3)}} = (5, 3)^T \quad \boldsymbol{x^{(4)}} = (7, 3)^T$$
$$\boldsymbol{x^{(5)}} = (1, 1)^T \quad \boldsymbol{x^{(6)}} = (3, 1)^T \quad \boldsymbol{x^{(7)}} = (5, 1)^T \quad \boldsymbol{x^{(8)}} = (7, 1)^T$$

Assume that the K-means algorithm is initialized with $\boldsymbol{\mu^{(1)} = (3.5, 1.5)^T}$ and $\boldsymbol{\mu^{(2)}} = (4.5, 1.5)^T$, i.e. we use $K = 2$ clusters.

3.1

Draw the data set in a coordinate system. Also draw $\boldsymbol{\mu^{(1)}}$ and $\boldsymbol{\mu^{(2)}}$

3.2

How does K-means assign data points to cluster centers? State for both $\boldsymbol{\mu^{(1)}}$ and $\boldsymbol{\mu^{(2)}}$ which data points are assigned to them

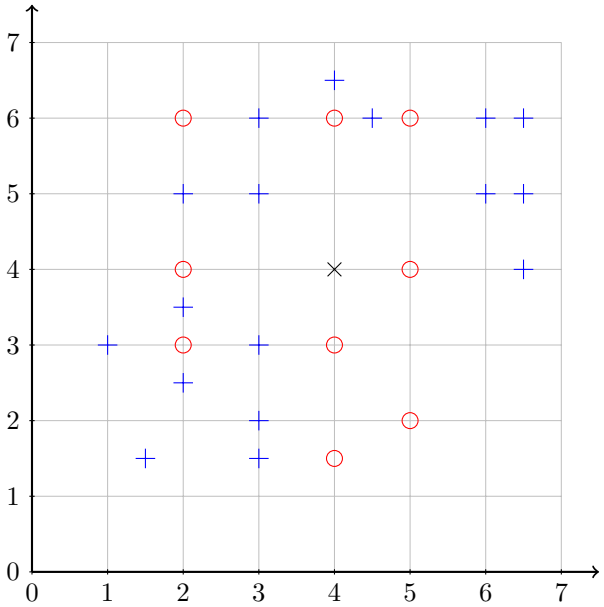
3.3

How does K-means update the cluster centers? Compute these updates for $\boldsymbol{\mu^{(1)}}$ and $\boldsymbol{\mu^{(2}}$

3.4

Compute the within-cluster squared distances *(k-means objective)* for the updated $\boldsymbol{\mu}^{(1)}$ and $\boldsymbol{\mu}^{(2}$. Will the objective change, if we perform another iteration of k-means? If yes, why? If no, why not?

4



Consider the training set for binary classification above, where + denotes the positive class (+1) and o the negative class (-1). Which class does the K-nearest neighbour classifier (with Euclidean distance) predict for the point marked with x, respectively for $K = 1$, $K = 3$, $K = 5$, and $K = 7$?

5

Explain the CART algorithm for learning decision trees. How is the decision tree initialized and how is it expanded? Give the names two impurity measures and explain how an impurity measure is used to expand the decision tree. What are typical stopping criteria for the learning algorithm?