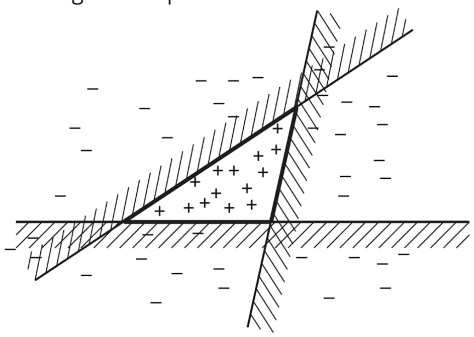
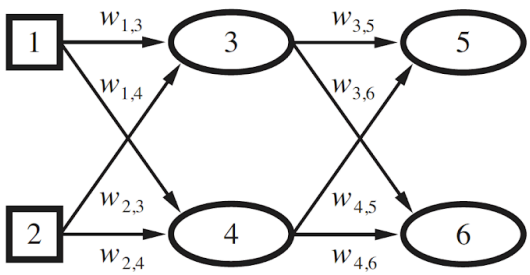
Questions | Chapter 18

1. The idea of ensemble learning methods is to select a collection, or ensemble, of hypotheses from the hypothesis space and combine their predictions, i.e. use multiple models. One example of such a technique is 'boosting'. The diagram below may be used to explain the idea behind ensemble learning. With the help of the diagram, explain how ensemble learning can be powerful.

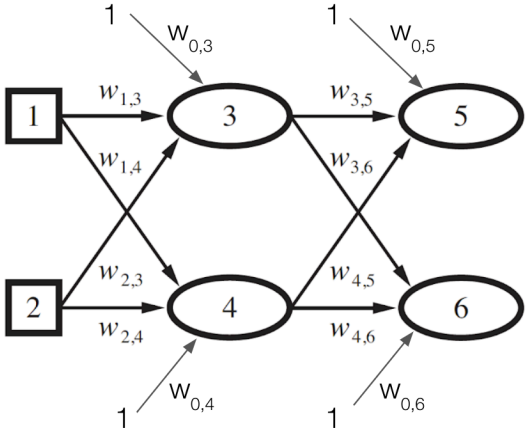


2. Below is a neural network with two inputs "1" and "2". Assuming that there are no biases in the neurons and that all neurons have linear activations, what will be the activation at node 5?



3. Below is a neural network with two inputs "1" and "2". Assuming that there are no biases in the neurons and that all neurons have logistic activations, what will be the activation at node 6?

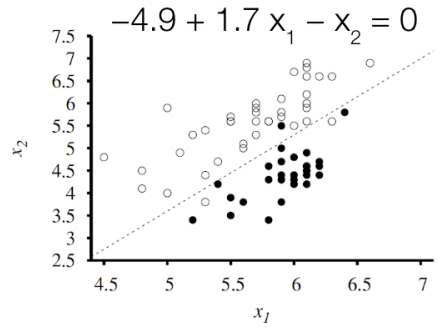
$$Logistic(z) = \frac{1}{1+e^{-z}}$$



4. What are the FOUR advantages of using a logistic function for classification?

$$Logistic(z) = \frac{1}{1+e^{-z}}$$

5. Linear functions can be used as classifiers using a hard threshold. Is the data below linearly separable? On a linearly separable data how will you use the linear equation to decide the class?



6. Learning involves adjusting weights to minimize a loss. Weight space is defined by all possible settings of the weights. Assume that we have a loss function "Loss(w)" that computes the loss on a given dataset and weights w. In general the search for best values of weights is a general optimization search problem in a continuous weights space. To minimize loss, we can use the gradient descent algorithm. Assuming that $\frac{\delta y}{\delta x}$ can be used to calculate the partial derivative of a function y = f(x), provide a pseudocode to obtain the optimal weights.

7. Suppose we have a logistic regression task at hand. The loss function $h_w(x)$ for input vector x is given by the following equation. The perceptron learning rule (gradient descent method) for updating weights is also provided to you. Derive the rule for updating weights for logistic regression, i.e. the expression for w_i.

$$h_{\mathbf{w}}(\mathbf{x}) = Logistic(\mathbf{w} \cdot \mathbf{x}) = \frac{1}{1 + e^{-\mathbf{w} \cdot \mathbf{x}}} \qquad w_i \leftarrow w_i - \alpha \frac{\partial}{\partial w_i} Loss(\mathbf{w})$$

8. Given the gradient descent rule for updating weights in a neural network, we need to decide the learning rate, alpha. As the learning process proceeds, can we increase alpha? Why?

$$w_i \leftarrow w_i - \alpha \frac{\partial}{\partial w_i} Loss(\mathbf{w})$$

9. Assume that we would like to perform univariate linear regression on a dataset consisting of M data points. If we would like to use the 'squared' loss function L_2, what will be the expression for Loss (h_w). h_w is the linear function (model) with weights w? Clearly define all the terms in your expression.

10. Assume that we would like to perform univariate linear regression on a dataset consisting of M data points. If we would like to use the 'absolute difference' loss function L_2, what will be the expression for Loss (h_w). h_w is the linear function (model) with weights w? Clearly define all the terms in your expression.

11. Define supervised, unsupervised, and reinforcement learning. Provide examples for each.

12. When developing machine learning applications, we usually focus on one class of learning problem - "from a collection of input–output pairs, learn a function that predicts the output for new inputs." What we learn (improvement) and how we learn (improvement techniques) depends on four factors. What are they?