

Problem Set 1 Answers

Jonathan McEntee

9/22/2018

2 Design a two-input perceptron that implements the boolean function $A \wedge \neg B$. Design a two-layer network of perceptrons that implements $A \oplus B$ (\oplus is XOR).

Imagine a two input perceptron with inputs A and B , weights w_A and w_B , and a threshold θ . Let $w_A = 1$, $w_B = -2$ and $\theta = \frac{1}{2}$. Then:

- if $A = 1$ and $B = 0$ then $w_A \cdot A + w_B \cdot B = 1 \cdot 1 + -2 \cdot 0 = 1 > \frac{1}{2} = \theta$. The perceptron outputs 1.
- if $A = 1$ and $B = 1$ then $w_A \cdot A + w_B \cdot B = 1 \cdot 1 + -2 \cdot 1 = -1 < \frac{1}{2} = \theta$. The perceptron outputs 0.
- if $A = 0$ and $B = 1$ then $w_A \cdot A + w_B \cdot B = 1 \cdot 0 + -2 \cdot 1 = -2 < \frac{1}{2} = \theta$. The perceptron outputs 0.
- if $A = 0$ and $B = 0$ then $w_A \cdot A + w_B \cdot B = 1 \cdot 0 + -2 \cdot 0 = 0 < \frac{1}{2} = \theta$. The perceptron outputs 0.

Therefore this perceptron implements $A \wedge \neg B$. If $w_A = 1$, $w_B = 1$ and $\theta = \frac{3}{2}$, the perceptron implements $A \wedge B$. If $w_A = 1$, $w_B = 1$, and $\theta = \frac{1}{2}$, the perceptron implements $A \vee B$.

Imagine a two layer perceptron network with inputs A and B . In the first layer consists of two perceptrons. One (C) implements $A \vee B$, and the second (D) implements $A \wedge B$ (henceforth D). Let the second layer's single perceptron (E) implement $C \wedge \neg D$. Then:

- if $A = 1$ and $B = 0$ then $C = 1$ and $D = 0$. Then $E = 1$
- if $A = 1$ and $B = 1$ then $C = 1$ and $D = 1$. Then $E = 0$
- if $A = 0$ and $B = 1$ then $C = 1$ and $D = 0$. Then $E = 1$
- if $A = 0$ and $B = 0$ then $C = 0$ and $D = 0$. Then $E = 0$

Therefore this perceptron network implements $A \oplus B$.

3 Derive the perceptron training rule and gradient descent training rule for a single unit with output $o = w_0 + w_1x_1 + w_2x_1^2 + \dots + w_nx_n^2$. What are the advantages of using gradient descent training rule for training neural networks over the perceptron training rule?

6 Imagine you had a learning problem with an instance space of points on the plane and a target function that you knew took the form of a line on the plane where all points on one side of the line are positive and all those on the other are negative. If you were constrained to only use decision tree or nearest-neighbor learning, which would you use? Why?

I would use a decision tree. K nearest neighbors holds a fundamental assumption that values which are closer to each other are more likely to be similar. This does not hold with a divided plane.

For example, imagine 4 points on a 2-D plane divided on the y-axis with the same y value and x values equal to -3, -2, -1, 1, 6, and 7. The points on the negative side of the y-axis are of one class and those on the positive side are on the other. A 3 nearest neighbor algorithm would misclassify -1 as a positive example. A decision tree however, could create branches for $x > 0$ or $x \leq 0$, which would perfectly classify the data.

The target concept will always be within the version space of the decision tree algorithm, but that is not the case with KNN.

7 Give the VC dimension of the following hypothesis spaces. Briefly explain your answers.

1. An origin-centered circle (2D)

Imagine a point $x \in \mathbb{R}^2$, and an origin-centered circle with radius r . I define x as being within the circle (+) if and only if $\|x\| \leq r$. Otherwise it will be outside of the circle (-). Therefore the VC dimension is greater than or equal to 1.

Imagine two points $x_1, x_2 \in \mathbb{R}^2$ such that $\|x_1\| \leq \|x_2\|$. For the VC dimension to be greater than or equal to 2, it must be possible for x_2 to be within the circle (+), while x_1 is outside of it (-). However x_2 being within the circle implies $\|x_2\| \leq r$, which in turn implies $\|x_1\| \leq r$. Therefore it is impossible to shatter x_1 and x_2 . The VC dimension is 1.

2. An origin-centered sphere (3D)

Like in 2-dimensional space, I consider a point $x \in \mathbb{R}^3$ inside the origin-centered sphere if and only if $\|x\| \leq r$. If you accept that definition, then following the same argument as before, the VC dimension is 1.