Machine Learning Exercises 4

Due: November 8

Tim van Erven

November 1, 2007

Exercises

1. The following Boolean functions take two Boolean features x_1 and x_2 as input. The features can take on the values -1 and $+1$, where -1 represents False and $+1$ represents True. The output y of the functions can also take on the values -1 and $+1$, with the same interpretation. For each of the functions below, either give weights for a perceptron such that the perceptron represents the function or argue that no such weights exist.

   Hint: Draw pictures like on slides 9 and 10 from mlslides8.pdf. (You do not have to submit these.)

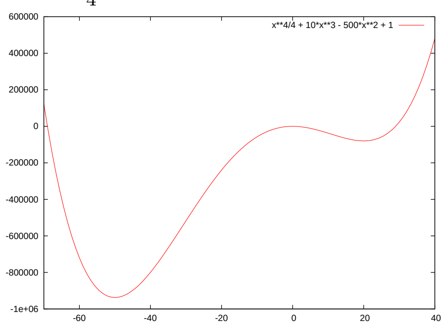
   (a) $y = \neg\text{AND}(x_1, x_2)$

   (b) $y = \begin{cases} +1 & \text{if } x_1 = x_2 \\ -1 & \text{otherwise} \end{cases}$

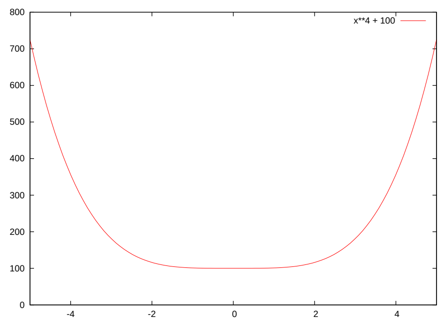
   (c) $y = \begin{cases} +1 & \text{if } x_1 = 1 \text{ and } x_2 = -1 \\ -1 & \text{otherwise} \end{cases}$

2. (a) For both of the following functions, argue whether gradient descent is an appropriate method to find the minimum.

   $\frac{1}{4}x^4 + 10x^3 - 500x^2 + 1$



   $x^4 + 100$



   (b) Suppose we run gradient descent for each of the functions, regardless of whether it is appropriate. What would be Δx_n for each of the functions when the learning rate is $\eta = 0.1$? (Work out the derivative.)

3. Suppose we have training data $D = \begin{pmatrix} y_1 \\ \mathbf{x}_1 \end{pmatrix}, \ldots, \begin{pmatrix} y_n \\ \mathbf{x}_n \end{pmatrix}$ and we want to use gradient descent to find weights $\mathbf{w}$ that minimize the error on D for a linear unit $h_{\mathbf{w}}$. However, instead of the Sum of Squared Errors (SSE), we use a strange new error measure called the Sum of Quadratic Errors (SQE). It is defined as

$$\text{SQE}(\mathbf{w}, D) = \sum_{i=1}^{n} (y_i - h_{\mathbf{w}}(\mathbf{x}_i))^4.$$

What would be the gradient that our algorithm would use in this case? Give a derivation like in Equation 4.6 of Mitchell.

Hints: See slides 28 and 29 of mlslides8.pdf, and Equation 4.6 in Mitchell. Note that Equation 4.6 applies the chain rule, so you may have to look that up somewhere.

Grading Policy

- Grades are between 1 and 10.
- You always start with 1 point.
- Partial points may be awarded for partially correct answers.