ECS 170: Problem Set 5

February 25, 2017

Your answers should be succinct - our solutions for each written problem are no more than a couple sentences.

1. What condition must hold on the training data such that the perceptron training algorithm can learn a classifier that makes perfect predictions on the training data?

2. Name two desirable properties of the linear unit objective function (shown below). D is the set of training data, t_d is the label of the dth training instance and $o_d = w_0 + w_1 x_1 + \ldots + w_m x_m$.

$$E[w] = \frac{1}{2} \sum_{d \in D} (t_d - o_d)^2$$

3. Name one possible way to avoid overfitting.

4. Give one major reason to use stochastic gradient descent over standard gradient descent.

5. Give an advantage of using sigmoid units over perceptron units, and an advantage of sigmoid units over linear units.

6. Construct a one-hidden-layer network of perceptrons that represents the XOR function of two inputs. For this problem, let `False` be encoded as -1 and `True` as 1.