Neural Networks Fall 2020

Gal Chechik, Shauli Ravofgel

Exercise 1

Due date: Novermber 2th

1. **(25 pts)** Given a set of samples $D = \{(x_i,y_i)\}_{i=1 \text{ to } N}$, we defined in class the problem of polynomial curve fitting as an error-minimization problem. Assuming that the labels $\{y_i\}_{i=1}$ N are i.i.d. and follow a Gaussian distribution whose mean is the prediction of the model $h_w(x)$ of the curve-fitting problem, $y_i \sim N(h_w(x_i),\sigma)$, show that maximizing the logarithm of the likelihood function is equivalent to minimizing the square error function $E(w) = \sum_{i=1 \text{ to } n} (h_w(x) - y_i)^2 / N$ we used in class. Recall that the likelihood is defined as $L = p(D|w)$.

2. **(25 pts)** We showed in class that for a zero-order polynomial (namely, a constant) the value that minimizes the squared error is the sample mean: $h_w(x) = \sum_{i=1 \text{ to } N} y_i/N$. Show that for the case of zero- order polynomial with an *absolute-value* error, $E(w) = \sum_{i=1 \text{ to } N} |h_w(x) - y_i|$, the optimal solution (the constant that minimizes the error) is the sample median.

3. **(25 pts)** The parity problem is defined as follows: given a set of *k* binary values (zeros and ones),

   output 1 if there is an odd number of 1's, and
   output 0 otherwise.

   a. Show that the perceptron algorithm discussed in class cannot learn to solve the parity

      problem, (hint,
      use k=2).

   b. Write an algorithm that uses one memory cell and can solve the parity problem efficiently.