Back Prop

User 910 • 3 Jun 2024

Constants

P: # of features N: size of training data M_i: size of ith hidden layer K: output size $y^{<i>}$: the ith training sample.

Perceptron

First, we talk a bit about the general setup of a neural network. A neural network is alternating linear function and non-linear function repeatedly. Any time a nonlinear operation happens indicates a new layer. So two things happen at every layer, a linear transformation, and a non-linear operation.

We start with the simplest setup, a neural network with no hidden layer, or a perceptron. We assume the output is of size K to be more general.

perceptron
perceptron