The Perceptron

1. Consider the function $f(x)=x^3+2x^2$ and the current solution $x_t=\langle 2\rangle$ compute one step of gradient descent with learning rate $\eta=0.1$.
2. Consider the function $f(x,y,z)=x^2+yz+yz^2$ and the current solution $x_t=\langle 1,1,-1\rangle$ compute one step of gradient descent with learning rate $\eta=0.1$.
3. The threshold of the activation function of the perceptron is a parameter to learn, as much as the weights. How can we put the threshold in the same form as the weights, so that the same update rule can be used for it?
4. What is the equation of the decision boundary for a perceptron? What does it represent?
5. Given the equation of the decision boundary of a perceptron in 2D (a straight line), what is the slope and what is the intercept?
6. What does it mean for two classes to be linearly separable in D dimensions?
7. Will the perceptron algorithm always converge to a point with zero error?
8. (Question for week 3) What error function do we use to derive the update rule for the perceptron? Why not the number of errors on a dataset? What is the advantage of the function we use over the number of errors?
9. Construct a perceptron able to separate the points: <1,1,0>, <2,3,1> where the last element is the class.
10. Add a new point to the dataset from the previous question, so that the perceptron you compute missclassifies it. Perform one step of gradient descent on the perceptron error to improve the error on the new data point.
11. Construct a perceptron able to separate the points <1,1,0>, <1, -1, 1>, and <-1, 1, 1> where the last element is the class.