Machine Intelligence 1 — **Exercise Sheet 02**
WS 2015/16, Obermayer/Lochmann — due: 29.10.2015

Connectionist Neurons and Multi Layer Perceptrons

2.1 Terminology (2 points)

(a) How does a nonlinear transfer function change the computational properties of a connectionist neuron? In which situations might this be useful?

(b) What is the function of the *bias* in a connectionist neuron? Give an example in which a classification with a `sign`$(\cdot)$ transfer function would *not* work without a bias (but would with one).

(c) What are *point* and *edge filters* and what are they used for?

(d) What is the difference between a *connectionist neuron* with a *logistic transfer function* and a *stochastic neuron*?

2.2 Finding Parameters of a Connectionist Neuron (5 points)

The dataset `applesOranges.csv` available on ISIS contains 200 measurements (`x.1` and `x.2`) from two types of objects as indicated by the column `y`. In this exercise, you should use a simple connectionist neuron with the sign function as transfer function to classify the objects i.e.

$$f(\mathbf{x}) = \mathrm{sgn}(\mathbf{w}^{\mathbf{T}}\mathbf{x} - \theta)$$

(a) Plot the data in a scatter plot (x.1 vs. x.2). Use color to indicate the type of each object.

(b) Set $\theta = 0$. Create a set of 19 equally spaced weight vectors $\mathbf{w} = [w_1, w_2]$ on the circle centered on $(0, 0)$ with radius 1. I.e. if α denotes the angle between the weight vector and the x-asis, for each weight $||w|| = 1$ and $\alpha_1 = 0, \alpha_2 = 10, ..., \alpha_{19} = 180$ such that $w_1 \in [-1, 1], w_2 \in [0, 1]$. For each weight vector $\mathbf{w}$ determine the classification performance ρ (% correct classifications) of the corresponding neuron and plot a curve showing α vs. ρ.

(c) From these weights, pick the weight vector yielding best performance. Now vary $\theta \in [-3, 3]$ and pick the value of θ giving the best performance.

(d) Plot the datapoints, colored according to the classification corresponding to these parameter values. Plot the weight vector w in the same plot. How do you interpret your results?

(e) Find the best combination of w and θ by exploring all combinations of α and θ.

2.3 Multi Layer Perceptrons (3 points)

(a) Describe a simple example in which a *multilayer perceptron* (MLP) can distinguish between two classes, but a single connectionist neuron can not.

(b) For a MLP with input $x \in \mathbb{R}$ and one hidden layer, the input-output function can be computed as

$$y(x) = \sum_{i=1}^{n_{\text{hid}}} w_i f(a_i(x - b_i))$$

with output weights w_i and parameters a_i and b_i for each hidden unit i. Create 50 MLPs with $n_{\text{hid}} = 10$ hidden units by sampling for each one a set of random parameters $\{w_i, a_i, b_i\}, i = 1, ..., 10$ and using $f := \tanh$ as the activation function. Use $a_i \sim \mathcal{N}(0, 2), w_i \sim \mathcal{N}(0, 1)$ and uniformly distributed $b_i \sim \mathcal{U}(-2, 2)$. Plot the input-output functions of these 50 MLPs for $x \in [-2, 2]$.

(c) Repeat this procedure using instead $a_i \sim \mathcal{N}(0, 0.5)$. What is the difference?

Bonus question: Compute the mean squared error between these 2x50 input-output functions and the function $g(x) = -x$. Which MLPs from these two classes approximate it best? Plot these 2 functions.

total:10