Multilayer neural networks
Can represent non-linear surface
No algorithm for learning in multi-layered
networks, and no convergence theorem!
Sigmoid activation function
0 = σ(w· x), where σ(z) = (1 / 1+ e^−z)
Forward propagation
Assume two-layer networks (i.e., one hidden layer):
I. For each test example:
1. Present input to the input layer.
2. Forward propagate the activations times the weights to each node in the hidden layer.
3. Forward propagate the activations times weights from the hidden layer to the output layer.
4. Interpret the output layer as a classification.
Momentum
The idea is to keep weight changes moving in the same direction
Δw^t = (learning rate) (δ error) (input) + (Momentum) (Δw^(t−1))
SVM
Instances are represented by vector x ∈ ℜn
input = traingin examples
output = alpha's, support vectors, bias
Margin
Distance from separating hyperplane to nearest positive (or negative) instance.
Length of margin (1 / ||w||)
Support Vectors
Elements of the training set that would changes the poistion of the hyperplane if removed.
Kernel function

3. Consider the following three points, $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$, which have been identified as support vectors for a training set. Here, y_i is the class of the point, and α_i is the suppo vector coefficient.

$\mathbf{x}_1 = (2, 1)$ $\quad y_1 = -1$ $\quad \alpha_1 = -4$

$\mathbf{x}_2 = (4, 3)$ $\quad y_2 = -1$ $\quad \alpha_2 = -4$

$\mathbf{x}_3 = (2, 3)$ $\quad y_3 = +1$ $\quad \alpha_3 = 8$

The bias is $b = 0$.

(a) Using the formula

$$h(\mathbf{x}) = \mathrm{sgn}\left(\left(\sum_{i=1}^{m} \alpha_i (\mathbf{x}_i \cdot \mathbf{x})\right) + b\right),$$

give the classification of the new instance $\mathbf{x} = (1, 2)$. Show your work.

$h(1,2) = sgn[-4(2,1)\cdot(1,2) - 4(4,3)\cdot(1,2) + 8(2,3)\cdot(1,2)]$

$= sgn[-4\cdot 4 - 4\cdot 10 + 8\cdot 8]$

$= sgn[-56 + 64] = sgn[8] = +1$

(b) Use the $\mathbf{x}_i$'s and α_i's to find the weight vector $\mathbf{w}$ associated with the separating hyperplane, where $\mathbf{w} = \sum_i \alpha_i \mathbf{x}_i$.

$W = -4(2,1) - 4(4,3) + 8(2,3)$

$= (-8,-4) + (-16,-12) + (16,24)$

$= (-8, 8)$

(c) Using the weight vector you obtained in part (b) and the bias b = 0, find the equation of the separating hyperplane. Give the equation in the slope-intercept form: $x_2 = (slope * x_1) + y\text{-}intercept$.

$-8x_1 + 8x_2 = 0$

$x_2 = x_1$