Assignment8 – Learning

Given: June 26 Due: July 1

Problem 8.1 (Decision List)

Construct a decision list to classify the data below. The tests should be as small as possible (in terms of attributes), breaking ties among tests with the same number of attributes by selecting the one that classifies the greatest number of examples correctly. If multiple tests have the same number of attributes and classify the same number of examples, then break the tie using attributes with lower index numbers (e.g., select A_1 over A_2).

Example	A_1	A_2	A_3	A_4	y
x_1	1	0	0	0	1
x_2	1	0	1	1	1
x_3	0	1	0	0	1
x_4	0	1	1	0	0

Problem 8.2 (General Properties of Linear Regression)

Consider a list of examples $(\vec{x}_i, y_i) \in \mathbb{R}^n \times \mathbb{R}$ for $i = 1, \ldots, m$. We want to apply linear regression.

1. What is the hypothesis space?
2. What is the point of the trick to set $(x_i)_0 = 1$ for all examples?
3. What is the maximum number of examples for which a consistent hypothesis can still exist?
4. How important is it to find a consistent hypothesis here?
5. What kind of loss function should we use here?

Problem 8.3 (Support Vectors)

Consider the following 2-dimensional dataset

support vector	*classification*
$\mathbf{x}_1 = \langle 0, 0 \rangle$	$\mathbf{y}_1 = -1$
$\mathbf{x}_2 = \langle 0, 0.5 \rangle$	$\mathbf{y}_2 = -1$
$\mathbf{x}_3 = \langle 0.5, 0 \rangle$	$\mathbf{y}_3 = -1$
$\mathbf{x}_4 = \langle 1, 1 \rangle$	$\mathbf{y}_4 = 1$
$\mathbf{x}_5 = \langle 2, 2 \rangle$	$\mathbf{y}_5 = -1$

1. Give a *linear separator* in the form $h(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b$ for the dataset containing only the examples for $\mathbf{x}_1$ to $\mathbf{x}_4$.
2. Explain informally why no linear separator exists for the full dataset of all 5 vectors.
3. Transform the dataset into a 3-dimensional dataset by applying the function $F(\langle u, v \rangle) = \langle u^2, v^2, u + v \rangle$.
4. Give a *linear separator* for the transformed full dataset in the form $h(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b$.

Problem 8.4 (Weight Updates)

We're trying to find a linear separator. Our examples are the set

Example number	$\mathbf{x}_1$	$\mathbf{x}_2$	y
1	2	0	2
2	3	1	2

Our hypothesis space contains the functions $h_{\mathbf{w}}(\mathbf{x}) = A(\mathbf{w}\cdot\mathbf{x})$ for 2+1-dimensional vectors $\mathbf{w}, \mathbf{x}$ (using the trick $\mathbf{x}_0 = 1$ to allow for the constant term $\mathbf{w}_0$) and some fixed function A.

As the initial weights, we use $\mathbf{w}_0 = \mathbf{w}_1 = \mathbf{w}_2 = 0$.

For each of the following cases, iterate the respective weight update rule once for each example (using the examples in the order listed). Use learning rate $\alpha = 1$.

1. Using the threshold function $A(z) = \mathcal{T}(z)$, i.e., $A(z) = 1$ if $z > 0$ and $A(z) = 0$ otherwise. Here we cannot do gradient descent, so we have to use the perceptron learning rule.
2. Using the logistic function $A(z) = 1/(1+e^{-x})$. Here we use gradient descent.