

CSCI567 Machine Learning (Fall 2007) Assignment #1

Instructor: Dr. Sofus A. Mackassy
TA: Cheol Han

Due time: 5:00pm, Sep 18, 2007

Student Name: _____

Student ID: _____

Score: _____

1. **(Logistic regression, 10 points)** In our definition of logistic regression, we defined the following equations:

$$P_1(x; w) = \frac{\exp w.x}{1 + \exp w.x}$$

$$P_0(x; w) = 1 - p_1(x; w)$$

Show that the logistic function is as follows:

$$P_1(x_i; w) = \frac{1}{(1 + \exp[-w.x_i])}$$

Using the above function, show that this equality holds:

$$\log \frac{P_1(x; w)}{P_0(x; w)} = w.x$$

2. **(Reject option, 20 points)** In many applications, the classifier is allowed to “reject” a test example rather than classifying it into one of the classes. Consider, for example, a case in which the cost of a misclassification is \$10 but the cost of having a human manually make the decision is only \$3, we can formulate this as the following loss matrix:

decision	true label y	
	0	1
predict 0	0	10
predict 1	10	0
reject	3	3

Suppose $P(y=1|x)$ is predicted to be 0.2. Which decision minimizes the expected loss? Now suppose $P(y=1|x)=0.4$, Now which decision minimizes the expected loss? Show that in cases such as this there will be two thresholds θ_0 and θ_1 such that the optimal decision is to predict 0 if $p_1 < \theta_0$, reject if $\theta_0 \leq p_1 \leq \theta_1$, and predict 1 if $p_1 > \theta_1$.

What are the values of θ_0 and θ_1 for the following loss matrix?

decision	true label y	
	0	1
predict 0	0	20
predict 1	15	0
reject	8	8

3. (**Hypothesis Spaces**, 30 points) Consider the hypothesis space of m -of- n functions. The m -of- n function states that if m out of n attributes are true, then predict +1. For example, consider the input space $\{x_1, x_2, x_3, x_4, x_5\}$; a specific m -of- n function would be: 1-of-3 of $\{x_1, x_3, x_5\}$, which means that if either x_1 , x_3 or x_5 are true in a given instance, then predict +1. In other words:

$$f(x_1, x_2, x_3, x_4, x_5)_{(1\text{-of-}3\{x_1, x_3, x_5\})} = \begin{cases} +1 & (x_1 \vee x_3 \vee x_5) = T \\ -1 & \text{otherwise} \end{cases}$$

A) What is the size of the m -of- n hypothesis space as a value of N , the number of boolean input values. In other words, how many possible hypotheses are there?

B) Which of these two is the more specific hypothesis and why?

1-of-3 of $\{x_1, x_2, x_4\}$

2-of-3 of $\{x_1, x_2, x_4, x_5\}$

C) When considering the parameter space of 5 boolean values $\{x_1, x_2, x_3, x_4, x_5\}$, what is an example of the most general (non-trivial). What is the most specific hypothesis?

4. (**Hypothesis Spaces**, 10 points) Imagine our hypothesis space is not one rectangle but a union of $m > 1$ rectangles.

A) What is the advantage of such a hypothesis class?

B) Show that class can be represented by such a hypothesis, given a large enough m .

C) If $f(\mathbf{x}) = +1$ if \mathbf{x} lies within one of the rectangles, how large would m need to be in the worst case?

5. (**Linear Threshold Units**, 30 points) Consider the points

$\langle 1, 2, +1 \rangle$

$\langle 1, 1, +1 \rangle$

$\langle 2, 4, -1 \rangle$

$\langle 2, 1, +1 \rangle$

$\langle 3, 4, -1 \rangle$

$\langle 3, 3, -1 \rangle$

A) Batch Perceptron Algorithm, learning rate $\eta = 1$, weight vector $\mathbf{w}_0 = \langle 0, 0, 0 \rangle$.

Calculate, by hand, the weight vectors ($\mathbf{w}_1, \mathbf{w}_2$) after the first two iterations. Show both the gradient vector and the weight vector.

B) Online Perceptron Algorithm, learning rate $\eta = 1$, weight vector $\mathbf{w}_0 = \langle 0, 0, 0 \rangle$.

Calculate, by hand, the final weight vector after having run through all the examples in the order given.

C) Batch Logistic Regression, learning rate $\eta = 1$, weight vector $\mathbf{w}_0 = \langle 0, 0, 0 \rangle$.

Calculate, by hand, the weight vectors ($\mathbf{w}_1, \mathbf{w}_2$) after the first two iterations. Show both the gradient vector and the weight vector.