CECS 551 Exam 1, Spring 2016, Dr. Ebert

Category C

1. Recall that there are several components to the process of converting data to knowledge. Name the two components that are most relevant to this course. Hint: model deployment is not one of them.

2. For a supervised-learning algorithm that returns a function $\hat{f} : X \to \{\texttt{true}, \texttt{false}\}$ that attempts to best approximate a target function $f : X \to \{\texttt{true}, \texttt{false}\}$, describe the primary required input to the algorithm.

3. Fill in the blank. Each data universe is composed of elements (or objects). Each element is composed of __________.

4. State the Inductive Learning Hypothesis.

5. State the addition axioms for a vector space V. Hint: assume $\overline{u}, \overline{v}$, and $\overline{w}$ are elements of V.

6. State the four fundamental properties of the dot product. Hint: assume $\overline{u}, \overline{v}$, and $\overline{w}$ are elements of dot product space V.

7. Given the linear equation $y = -3x + 7$, provide a vector $\overline{w}$ that is normal to this line.

8. Assume $\overline{c}_+$ is the mean of the positve data samples, $\overline{c}_-$ is the mean of the negtive data samples, and $\overline{c}$ is the midpoint of $\overline{c}_+$ and $\overline{c}_-$, then provide an equation for the decision surface that is constructed by the Simple Learning Algorithm.

9. State one desirable and one undesirable property of the Perceptron Learning algorithm, in relation to other algorithms that may be applied to a linearly-separable training set.

10. What does it mean to be a maximum-margin linear classifier? a support vector of such a classifier?

Category B

B1. For what values of a will the following system have i) no solution? ii) infinitely many solutions? iii) exactly one solution? Hint: transform into reduced row-echelon form.

$$x + 2y - 3z = 4$$

$$3x - y + 5z = 2$$

$$4x + y + (a^2 - 14)z = a + 2$$

B2. If the positively classified training vectors are

$$(-1,1),(1,3),(2,2),(3,2),$$

while the negatively classified training vectors are

$$(-2,-4),(1,-1), \text{ and } (4,-2),$$

then compute an optimal decision line that separates this data. Compute w and b for this line.

B3. If the positively classified training vectors are

$$(-1, 1), (1, 3), (2, 2), (3, 2),$$

while the negatively classified training vectors are

$$(-2, -4), (1, -1), \text{ and } (4, -2),$$

then convert these vectors to positive vectors and demonstrate the Perceptron Learning algorithm with $\eta = 1$ on these positive vectors. Start with $\overline{w}_0 = \overline{0}$, and use the order (with these vectors replaced by their positive equivalents)

$$(3, 2), (4, -2), (2, 2), (-2, -4), (-1, 1), (1, -1), (1, 3)$$

when checking for misclassifications.

Category A

A1. Given vectors $\overline{u}$ and $\overline{v}$ of a dot product space, prove the **parallelogram equality**

$$|\overline{u}+\overline{v}|^2 + |\overline{u}-\overline{v}|^2 = 2(|\overline{u}|^2 + |\overline{v}|^2).$$

Hint: use properties of the dot product.

A2. Suppose two parallel hyperplanes P_1 and P_2 have respective equations $\overline{w}\cdot\overline{x} = b_1$ and $\overline{w}\cdot\overline{x} = b_2$, where $\overline{w} \in \mathcal{R}^n$ and $b_1, b_2 \geq 0$ are constants. Provide a formula that gives the distance $d(P_1, P_2)$ between the two planes. Prove that your formula is correct.

A3. Consider the Perceptron Learning (with $\eta = 1$) where the input training vectors are all positive, and the algorithm is computing a decision surface that passes through the origin (i.e. $b = 0$) . Let r denote the maximum length of any training vector, and let $\overline{w}_k$, $k \geq 0$, denote the k th estimation of the vector that is normal to the decision surface. Assuming $\overline{w}_0 = \overline{0}$, use induction on k to prove that $|\overline{w}_k|^2 \leq kr^2$. Hint: use the Cauchy-Schwarz inequality.