UNIVERSITY OF WISCONSIN-LA CROSSE
Department of Computer Science

CS 419/519 **Machine Learning** **Fall 2018**
Midterm Exam (Practice) **24 October 2018**

- **Do not turn the page until instructed to do so.**
- **This is an open-book, open-notes exam.**
- **This booklet contains 6 pages including the cover page.**
- **You have 55 minutes.**
- **The maximum possible is 40 points.**
- **Write all your answers on the exam sheets.**

PROBLEM	SCORE
1	8
2	10
3	10
4	12
TOTAL	40

Answer Key

1. **(8 pts.)** **SHORT ANSWER PROBLEMS.**

Circle the appropriate answer and fill in the blanks where required.

a. (*4 pts.*) Data-set **A** (end of exam) **(** $\boxed{\textbf{IS}}$ **/ IS NOT)** linearly separable, because choosing, for example, the point 0.4 as the separator places all elements of class X below the separator, and every element of class Y at or above the separator.

Data-set **B** (end of exam) **(IS /** $\boxed{\textbf{IS NOT}}$**)** linearly separable, because any point chosen to place every X element below it must also put at least data-point $(0.2, Y)$ on the wrong side. Similarly, any point that places every Y element at or above it must place at least two X elements on the wrong side.

b. (*4 pts.*) If we use a radial basis function kernel on 2-dimensional data, the result is a translation of that data to three (3) dimensions.

Such a function is of the form: $k(\mathbf{x}, \mathbf{z}) = e^{-\frac{||\mathbf{x}-\mathbf{z}||^2}{2\sigma^2}}$.

It has highest value at point $\mathbf{z}$ itself and that highest value is equal to one (1.0). The parameter σ controls the diameter at which the function drops to 0, moving away from $\mathbf{z}$.

2. **(10 pts.)** **DECISION TREES.**

(a) (5 pts.) Compute the information-theoretic entropy of data-set **C** (end of exam). Show all necessary work; results should be accurate to no less than 3 decimal places.

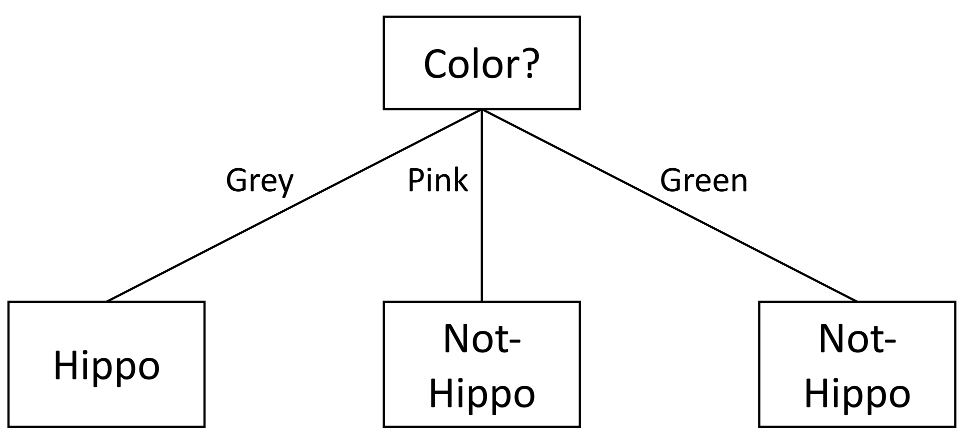
If we let $H \equiv Hippo$ and $\neg H \equiv$ *Not-Hippo*, we have:

$$P(H) = \frac{H}{H + \neg H} = \frac{3}{5} = 0.6 \qquad P(\neg H) = \frac{\neg H}{H + \neg H} = \frac{3}{5} = 0.4$$

Thus, we have:

$$\begin{aligned} H(\mathbf{C}) &= -(P(H)\log_2 P(H) + P(\neg H)\log_2 P(\neg H)) \\ &= -(0.6 \log_2 0.6 + 0.4 \log_2 0.4) \\ &= -(-0.442 + -0.529) \\ &= 0.971 \text{ bits} \end{aligned}$$

(b) (5 pts.) Draw the decision tree that the algorithm covered in class would produce (using information gain as the principle for choosing attributes on which to split data).



3. **(10 pts.)** **CLASSIFICATION METHODS.**

Suppose we have a 2-dimensional data point, belonging to the class 0: $(\mathbf{x}, y) = (0.1, 0.2, 0)$. Suppose further that we have a weight vector: $(1, 1, 1)$

(a) (4 pts.) Compute the output of the logistic function upon this data-point, using that weight vector. Show all necessary work; results should be accurate to no less than 3 decimal places.

We have $\mathbf{w} \cdot \mathbf{x} = (1 + 0.1 + 0.2) = 1.3$, and so:

$$h_{\mathbf{w}}(\mathbf{x}) \;=\; \frac{1}{1 + e^{-\mathbf{w} \cdot \mathbf{x}}} \;=\; \frac{1}{1 + e^{-1.3}} \;=\; \frac{1}{1.273} \;=\; 0.786$$

(b) (6 pts.) Compute the new weight vector we will get after doing a single iteration of logistic regression updating based upon the error made on the data-point given. Again, show all work, and use no less than 3 digits of decimal accuracy.

The logistic update equation for weights is:

$$w_j \;\leftarrow\; w_j + \alpha(y_i - h_{\mathbf{w}}(\mathbf{x})) \times h_{\mathbf{w}}(\mathbf{x})(1 - h_{\mathbf{w}}(\mathbf{x})) \times x_{i,j}$$

and in this case, the error made is:

$$(y_i - h_{\mathbf{w}}(\mathbf{x})) \;=\; (0 - 0.786) \;=\; -0.786$$

and the logistic function derivative term is:

$$h_{\mathbf{w}}(\mathbf{x})(1 - h_{\mathbf{w}}(\mathbf{x})) \;=\; 0.786\,(1 - 0.786) \;=\; 0.786 \times 0.214 \;=\; 0.168$$

giving us the product (assuming parameter $\alpha = 1$):

$$\alpha(y_i - h_{\mathbf{w}}(\mathbf{x})) \times h_{\mathbf{w}}(\mathbf{x})(1 - h_{\mathbf{w}}(\mathbf{x})) \;=\; -0.786 \times 0.168 \;=\; -0.132$$

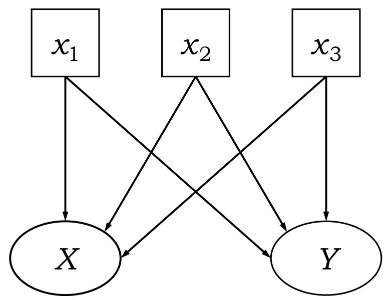
and so we do updates as follows (assuming parameter $\alpha = 1$:

$$\begin{aligned}
w_0 &\leftarrow 1 + (-0.132 \times 1) \;=\; 1 - 0.132 \;=\; 0.868 \\
w_1 &\leftarrow 1 + (-0.132 \times 0.1) \;=\; 1 - 0.0132 \;=\; 0.9868 \\
w_2 &\leftarrow 1 + (-0.132 \times 0.2) \;=\; 1 - 0.0264 \;=\; 0.9736
\end{aligned}$$

4. **(12 pts.)** **NEURAL NETWORKS.**

(a) (4 pts.) Draw a single-layer perceptron network for data-set **D** (end of exam). Assume the two layers are fully inter-connected. Label the output neurons with the relevant classes for which they stand in your network.

Note: for this question, if you had the bias-weights for the output neurons represented using an explicit arrow and/or node, that is fine; either way is correct here.



(b) (8 pts.) Assuming all weights (including bias weights) are set initially to 0.1, compute the output of each output neuron in the network, and the error they make, on the single data-point: $(\mathbf{x}, \mathbf{y}) = (0.5, 0.4, 0.2, (0, 1))$. Show all necessary work; results should use no less than 3 digits of decimal accuracy.

Since all weights are the same, the input function is the same of both output neurons:

$$in_j \;=\; (0.1 + (0.1 \times 0.5) + (0.1 \times 0.4) + (0.1 \times 0.2) \;=\; (0.1 + 0.05 + 0.04 + 0.02) \;=\; 0.21$$

Similarly, both neurons produce the same output (using the logistic activation function):

$$g(in_j) \;=\; \frac{1}{1 + e^{-in_j}} \;=\; \frac{1}{1 + e^{-0.21}} \;=\; \frac{1}{1.811} \;=\; 0.552$$

Therefore, we have errors for each output neuron as follows:

$$\begin{aligned} Err_X &= (0 - 0.552) = -0.552 \\ Err_Y &= (1 - 0.552) = 0.448 \end{aligned}$$

Data-set A (question 1): 1-dimensional numerical data-points belonging to two classes, X and Y:

$$\{(0.1, X),\ (0.2, X),\ (0.3, X),\ (0.4, Y),\ (0.5, Y),\ (0.6, Y)\}$$

Data-set B (question 1): 1-dimensional numerical data-points belonging to two classes, X and Y:

$$\{(0.1, X),\ (0.2, Y),\ (0.3, X),\ (0.4, X),\ (0.5, Y),\ (0.6, Y)\}$$

Data-set C (question 2): 2-dimensional data-points, each with two features, Color and Size, belonging to two classes, *Hippo* and *Not-Hippo*:

$$\{\ (Grey, Large, Hippo),\ (Grey, Medium, Hippo),\ (Grey, Small, Hippo),\\ (Pink, Small, \textit{Not-Hippo}),\ (Green, Large, \textit{Not-Hippo})\ \}$$

Data-set D (question 4): 3-dimensional numerical data-points belonging to two classes, X and Y; in this data-set an output vector $\mathbf{y} = (1, 0)$ means the data is of type X, while vector $\mathbf{y} = (0, 1)$ means the data is of type Y:

$$\{\ (0.5, 0.4, 0.2, (0, 1)),\ (0.1, 0.2, 0.3, (0, 1)),\ (0.3, 0.4, 0.5, (0, 1)),\\ (0.9, 0.8, 0.5, (1, 0)),\ (0.8, 0.9, 0.5, (1, 0)),\ (0.9, 0.9, 0.4, (1, 0))\ \}$$