CMSC422: Practice Problems 2

1 Naïve Bayes Classifiers

Consider three random variables X_1, X_2 and Y. X_1 and Y are generated according to Tables 1 and 2. Once X_1 has been generated, X_2 is assigned the same value as X_1.

	$P(Y)$
$Y = 0$	0.7
$Y = 1$	0.3

Table 1: $P(Y)$

	$X_1 = 0$	$X_1 = 1$
$Y = 0$	0.6	0.4
$Y = 1$	0.4	0.6

Table 2: $P(X_1|Y)$

Consider a Naïve Bayes classifier trained on features X_1 and X_2 to predict class label Y, using a very large number of training examples generated from the probability distributions in Tables 1 and 2.

1. Express $P(X_2 = x|Y = y)$ as a function of probabilities given in Tables 1 and 2, and use the resulting formula to fill up Table 3.

	$X_2 = 0$	$X_2 = 1$
$Y = 0$		
$Y = 1$		

Table 3: $P(X_2|Y)$

2. What is the prediction y_a of the classifier for $X_1 = 0$ and $X_2 = 0$? Explain.
3. What is the prediction y_b of the classifier for $X_1 = 0$ and $X_2 = 1$? Explain.
4. What is the prediction y_c of the classifier for $X_1 = 1$ and $X_2 = 0$? Explain.
5. What is the prediction y_d of the classifier for $X_1 = 1$ and $X_2 = 1$? Explain.

Now consider a second naïve Bayes classifier trained without the duplicate feature X_2.

1. What predictions does the second classifier make? Do they agree with those of the first classifier?

2. What is the impact of the duplicate feature X_2 on the probabilities of the predictions made by the 1st classifier compares to the 2nd?

2 Expressiveness of Neural Networks

Consider neural networks built out of units that take real-valued input $X_1, ..., X_n$ where the unit output Y is given by
$Y = \frac{1}{1+exp(-(w_0+\sum_i w_i X_i))}$.
In this problem, we explore the ability of such neural networks to represent boolean functions. The inputs X_i will be 0 or 1. The output Y will be real-valued, ranging between 0 and 1. We will interpret Y as a boolean value by interpreting it to be a boolean 1 if $Y > 0.5$, and interpreting it to be a boolean 0 otherwise.

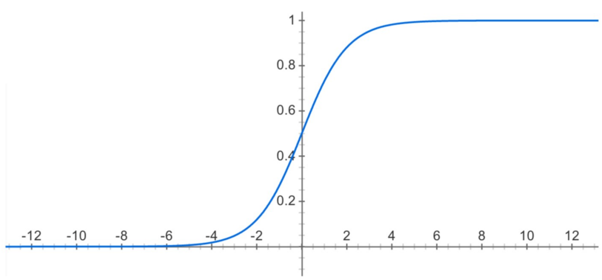

Figure 1: Non-linearity function $f(x) = \frac{1}{1+exp(-x)}$

- Using Figure 1, give 3 weights for a single unit with inputs X_1 and X_2 that implements the logical OR function $Y = X_1 \vee X_2$.
- Give 3 weights for a single unit with inputs X_1 and X_2 that implements the logical AND function $Y = X_1 \wedge X_2$.
- It is impossible to implement XOR ($Y = X_1 \oplus X_2$) in a single unit. However, you can do it with a neural network of multiple units. Design a network to implement XOR with the smallest number of unit you can. Draw it and show the weights for each unit.
- Create a network with one hidden layer that implements the function $Y = (A \vee \neg B) \oplus (\neg C \vee \neg D)$ Draw your network, and show all weights of each unit.

3 Bias Updates in 2-Layer Feedforward Neural Network

Consider the neural network defined in Figure **??**. We want to train it using the backpropagation algorithm to minimize squared error on the training set.

Recall that derived update rules for a similar neural network without explicit bias terms in class.

1. Derive the update rule for parameters v_{b_h}.
2. Derive the update rule for parameters w_{1,b_i} and w_{2,b_i}.
3. The bias is sometimes referred to as a pseudo-feature. Explain why, based on your answers to the two previous questions.