CSE 575: Homework #2

Due: October 24, 2022

Problem 1

Consider the following data points, also plotted in Figure 1:

$$\text{Labeled } +1 : (-2, 0), (0, 2)$$
$$\text{Labeled } -1 : (2, 2), (2, 0), (3, -1).$$

The positive data points are represented as blue circles, and the negative points as red triangles. Given this data, we wish to train a hard-margin linear SVM classifier.

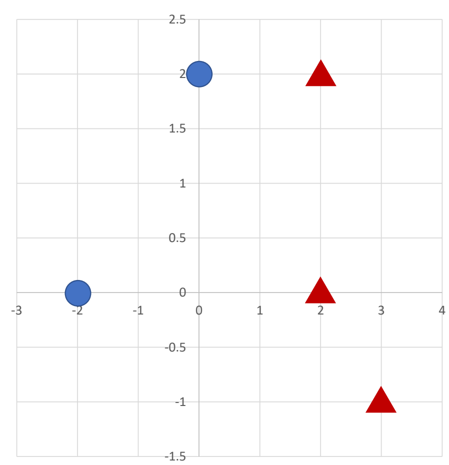


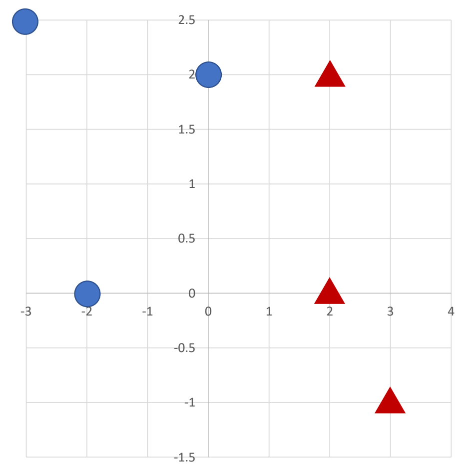
Figure 1: Data points

a) (6pt) What is the optimization problem for the maximum margin classifier (also called the *primal* problem) on this dataset? Denote the optimization variables by w_1, w_2, b.

b) (6pt) What is the dual formulation of the problem in part (a)? Denote the optimization variables by a_1, a_2, a_3, a_4, a_5

c) (6pt) What is the decision boundary of the SVM classifier? *Hint: You may make a geometric argument using* ***figure****. You are not required to solve the optimization problem.*

d) (6pt) Suppose we add a new data point at $(-3, 2.5)$ (as shown in the figure below). How will the decision boundary change in this case?



Problem 2

Figure 2 plots decision boundaries of SVM classifiers using different kernels and/or different slack penalty C. The data points labeled $+1$ and -1 are represented by circles and triangles, respectively. The support vectors for each decision boundary is illustrated as solid circles and triangles. For each of the following scenarios, find the matching plot from Figure 2 (there is a one-to-one match). Justify each choice briefly in 1–2 sentences.

a) (5pt) A hard-margin linear SVM.

b) (5pt) A soft-margin linear SVM with $C = 0.1$.

c) (5pt) A soft-margin linear SVM with $C = 10$.

d) (5pt) A hard-margin kernel SVM with $k(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T\mathbf{z})^2$.

e) (5pt) A hard-margin kernel SVM with $k(\mathbf{x}, \mathbf{z}) = \exp(-\|\mathbf{x} - \mathbf{z}\|^2)$.

Problem 3

Recall the perceptron $y(\mathbf{x}) = f(\mathbf{w}^T\mathbf{x} + b)$ where $f(a)$ is the step function:

$$f(a) = \begin{cases} 1 & \text{if } a \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

In this question, you will construct a multilayer perceptron with a single hidden layer to represent the XNOR function.

a) (8pt) Consider a perceptron with a two-dimensional input: $y(x_1, x_2) = f(w_1x_1 + w_2x_2 + b)$. What is the value of w_1, w_2, b such that this perceptron represents the logical conjunction (AND) between x_1 and x_2? Assume that the inputs are Boolean: $x_1, x_2 \in \{0, 1\}$.

b) (8pt) Again, consider a perceptron with two Boolean inputs: $y(x_1, x_2) = f(w_1x_1 + w_2x_2 + b)$. What is the value of w_1, w_2, b such that this perceptron represents the logical NOR between x_1 and x_2? In other words, the function should output 1 if both inputs are 0, and output 0 otherwise.

Problem 3 continued on next page. . .

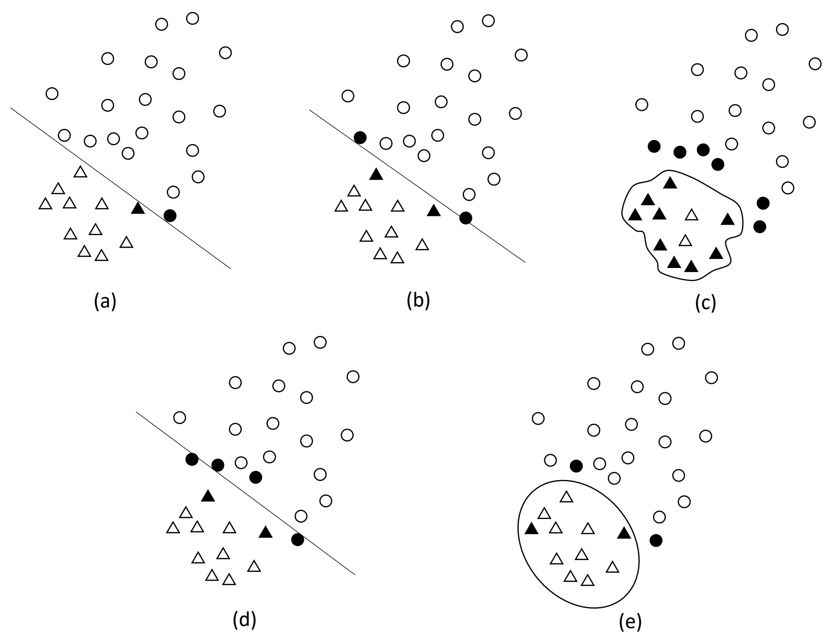
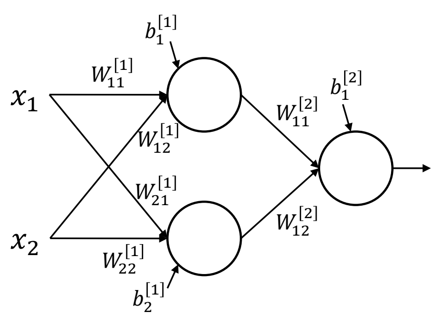


Figure 2: SVM decision boundaries

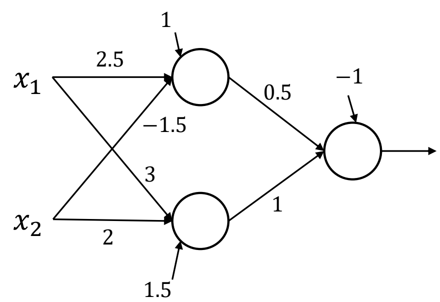
c) (10pt) The logical exclusive nor (XNOR) between x_1 and x_2 is 1 if the inputs are equal (both 0 or both 1); and 0 otherwise. **(i)** Show why XNOR cannot be represented by a single perceptron. You may use a geometric argument with a simple figure (the figure may be a hand-drawn image). **(ii)** Derive the weights for the following multilayer perceptron such that it outputs the XNOR between x_1 and x_2. *Hint: can you write XNOR using AND, NOR, and OR?*



Problem 4

The figure below shows a feedforward neural network with a single two-unit hidden layer and an output unit. Assume all neurons use the sigmoid activation function. Let $\mathbf{x} = (x_1, x_2) = (0.5, 0.1)$ be an example with label 1.

a) (8pt) By forward propagation, calculate the output values of the neurons for the input $\mathbf{x}$. In other words, what are the values of $z_1^{[1]}, z_2^{[1]}, z_1^{[2]}$?



b) (10pt) Suppose the error function is the squared error: $E = \frac{1}{2}(y(\mathbf{x}) - t)^2$ where t is the label for $\mathbf{x}$ and $y(\mathbf{x})$ the output of the neural network given $\mathbf{x}$. By backpropagation, calculate the following partial derivatives (note that $W_{ij}^{[l]}$ denotes the weight associated with the j-th input of the i-th node in layer l):

$$\frac{\partial E}{\partial W_{11}^{[1]}}, \quad \frac{\partial E}{\partial W_{12}^{[1]}}, \quad \frac{\partial E}{\partial W_{21}^{[1]}}, \quad \frac{\partial E}{\partial W_{22}^{[1]}}, \quad \frac{\partial E}{\partial W_{11}^{[2]}}, \quad \frac{\partial E}{\partial W_{12}^{[2]}}$$

c) (7pt) What will be the weights after a *single* step of gradient descent using the example $\mathbf{x}$ with a learning rate of $\eta = 0.1$?