CS 165B: Homework #3

Due on May 20, 2023

Xifeng Yan 07575

Qiyuan Zhuang

Problem 1

(a) **What is the corresponding feature mapping space? Express your answer in vector form in terms of x.**

$\kappa(\boldsymbol{x},\boldsymbol{z}) = (x_1z_1 + x_2z_2 + 2)^2 = x_1^2z_1^2 + x_2^2z_2^2 + 2x_1z_1x_2z_2 + 4x_1z_1 + 4x_2z_2 + 4$
feature mapping space: $\{\boldsymbol{y} = [y_1, y_2, y_3, y_4, y_5, 2]^T = \varphi(\boldsymbol{x}) = [x_1^2, x_2^2, \sqrt{2}x_1x_2, 2x_1, 2x_2, 2]^T\}$

(b) **Show that this feature mapping space has a dot product which is equivalent to the original kernel function.**

$$\begin{aligned}\boldsymbol{y}\cdot\boldsymbol{y'} &= \sum_{i=1}^{5} y_iy_i' + 4\\ &= x_1^2x_1'^2 + x_2^2x_2'^2 + 2x_1x_1'x_2x_2' + 4x_1x_1' + 4x_2x_2' + 4\\ &= (x_1x_1' + x_2x_2' + 2)^2\\ &= \kappa(\boldsymbol{x},\boldsymbol{x'})\end{aligned}$$

Problem 2

1. **k = 3 and simple majority vote?**
   class B. Picked among d_1, d_2, d_3.

2. **k = 5 and simple majority vote?**
   class A. Picked among d_1, d_2, d_3, d_4, d_5.

3. **k = 3 and weighted by cosine similarity? This means each vote is weighted by its cosine similarity to d.**
   class B. A : 1 B : 0.95+0.94=1.89

4. **k = 5 and weighted by cosine similarity?**
   class B. A : 1+0.45+0.4=1.85 B : 0.95+0.94=1.89

Problem 3

Please show how w will be updated using the Perceptron algorithm with learning rate $\eta = 0.5$. Suppose that there is no bias. You only need to show the results for one iteration on all the data samples, i.e., only need to show five updates on w here.

update 1: $\mathbf{w}_1 = (1,1,1) - 0.5\mathbf{x}_1 = (2, -0.5, 0.5)$
update 2: $y_2\mathbf{w}_1\cdot\mathbf{x}_2 = 8.5 > 0$, thus no update here. $\mathbf{w_2} = \mathbf{w_1}$
update 3: $\mathbf{w}_3 = \mathbf{w}_2 - 0.5\mathbf{x}_3 = (0.5, -3, 0)$
update 4: $\mathbf{w}_4 = \mathbf{w}_3 + 0.5\mathbf{x}_4 = (1.5, -2.5, 1.5)$
update 5: $y_5\mathbf{w}_4\cdot\mathbf{x}_5 = 5.5 > 0$, thus no update here. $\mathbf{w_5} = \mathbf{w_4} = (\mathbf{1.5}, \mathbf{-2.5}, \mathbf{1.5})$

Problem 4

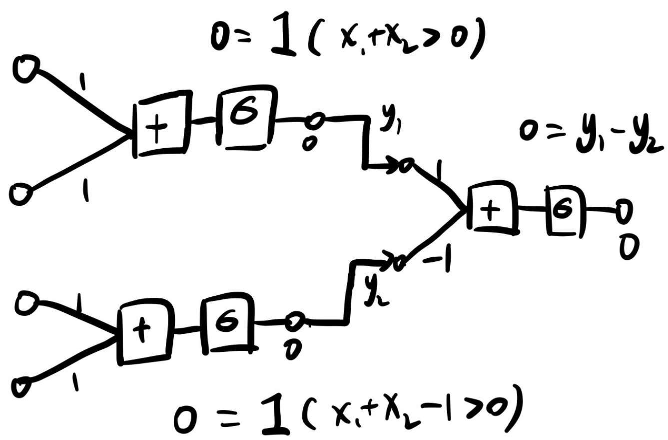
Perceptron with XOR data

(a) **Explain why a Perceptron cannot classify this XOR dataset.**

The dataset is not linearly separable, which means we cannot draw a straight line to seperate the data of classes. If we use Perceptron, it will go into an infinite loop.

(b) **If a Perceptron cannot classify XOR data, is it possible to stack two Perceptrons together so that it can classify XOR data correctly? If yes, draw the network; if not, please explain the reason.**

Yes, and we can implement two-level Perceptrons to reach the goal. For the first level, we can use two Perceptrons, one of AND function, the other of OR function. In the second level, we can use single Perceptron to classify.



(c) **Show a possible solution to handle XOR data (using only one Perceptron)**
We can use transformation first:

$$(0,0)->(0,0,0)$$

$$(0,1)->(0,1,1)$$

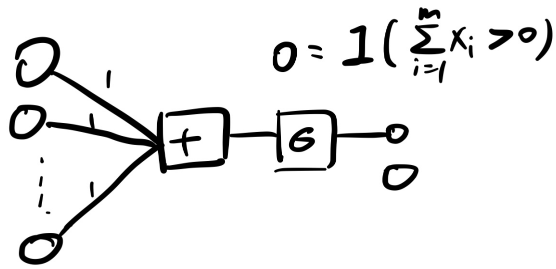
$$(1,0)->(1,0,1)$$

$$(1,1)->(1,1,0)$$

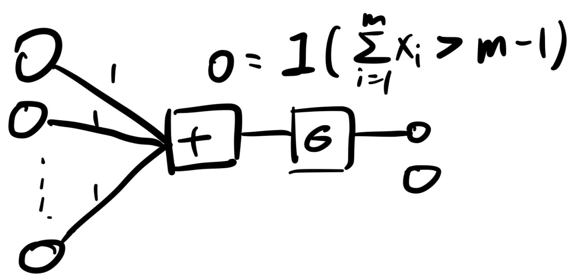
Then we can perform perceptron: $(0,0,0)$ and $(1,1,0)$ map to 1. $(0,1,1)$ and $(1,0,1)$ map to 0.

Problem 5

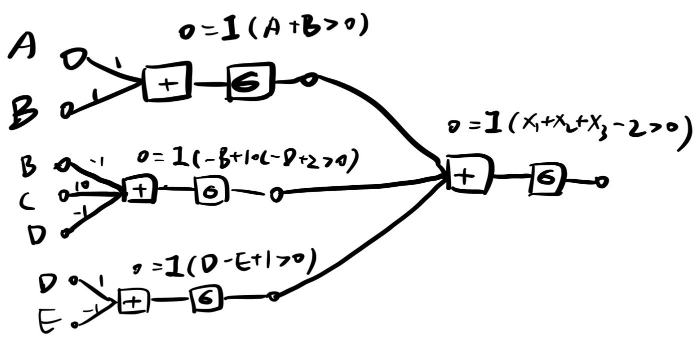
1. **Compute the OR function of m inputs**



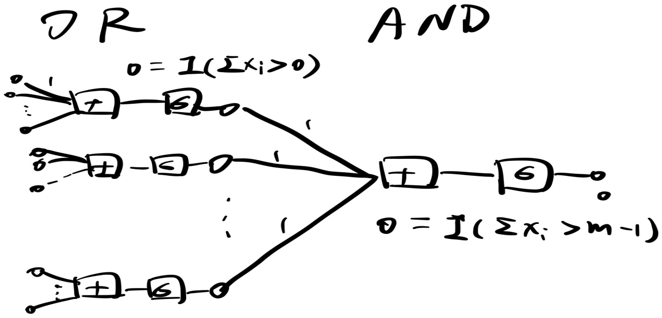
2. **Compute the AND function of n inputs**



3. **2-stacked perceptrons for computing the function**



4. **2-stacked perceptrons to compute any (given) logical expression**



Problem 6

(a) **Intuitively, where could this data point lie relative to (H0 , H1, H2) when $\xi_i = 0$? Is this data point classified correctly?**
H0. Yes.

(b) **Where could this data point lie relative to (H0 , H1, H2) when $0 < \xi_i < 1$? Is this data point classified correctly?**
Between H0 and H1. Yes.

(c) **Where does this data point lie relative to (H0 , H1, H2) when $\xi_i = 1$?**
H1.

(d) **Finally, where does it lie relative to (H0 , H1, H2) when $\xi_i > 1$? Is this data point classified correctly?**
Below H1. No.

Problem 7

Find the optimal linear support vector machine weights and equation that separates the categories by maximizing the margin. Also write the equation for the H+ and H- hyperplanes.

Since there's only one negative point, we can solve the problem simply. The line connecting two positive points: $x_1 + x_2 - 1.5 = 0$. If we translate it such that it will cross the negative poin, it'll be: $x_1 + x_2 - 2.5 = 0$. Then we obtain the correct decision boundary:

$$x_1 + x_2 - 2 = 0$$

in which weights are 1,1.
Additionally, H+: $x_1 + x_2 - 1.5 = 0$, H-: $x_1 + x_2 - 2.5 = 0$.

Problem 8

(a) **Show that equation (1) is equivalent to the following problem.**

For equation (1), we minimize m choosing i, then we minimize $\|w\|_2$. If we set $m' = km$, then margin $= \frac{km}{k\|w\|_2}$. We can choose a k such that $km = 1$,which means subject to $m \geq 1$. We still need to minimize $\|w\|$, also equivalent to min $\frac{1}{2}\|w\|_2^2$.

(b) **Explain how C affects the margin and the margin error.**

C controls tradeoff between the size of the margin and the amount of misclassification.
A large C places a high penalty on misclassified points because $C \sum \xi_i$ will be large and thus request ξ_i to be small.It leads to a small margin(margin error). A smaller C places less matters on correctly classifying of the data points, maximizing the margin. When $C = 0$, the SVM doesn't care about any misclassification, but only to maximize the margin as well as margin error.