

CS246: Mining Massive Datasets

Homework 4

Answer to Question 1(a)

| w | w_1 | w_2 | w_3 |
|--------------------------------|-------|-------|-------|
| Initial | 0 | 0 | 0 |
| After Observing Patient ID = 1 | 0 | 0 | -1/5 |
| After Observing Patient ID = 2 | 0 | 0 | -1/5 |
| After Observing Patient ID = 3 | 0 | 0 | -1/5 |
| After Observing Patient ID = 4 | 0 | -1/5 | 0 |
| After Observing Patient ID = 5 | -1/5 | -1/5 | 1/5 |

Answer to Question 1(b)

The Perceptron algorithm will not return a solution to those data. Those data points are not linearly separable.

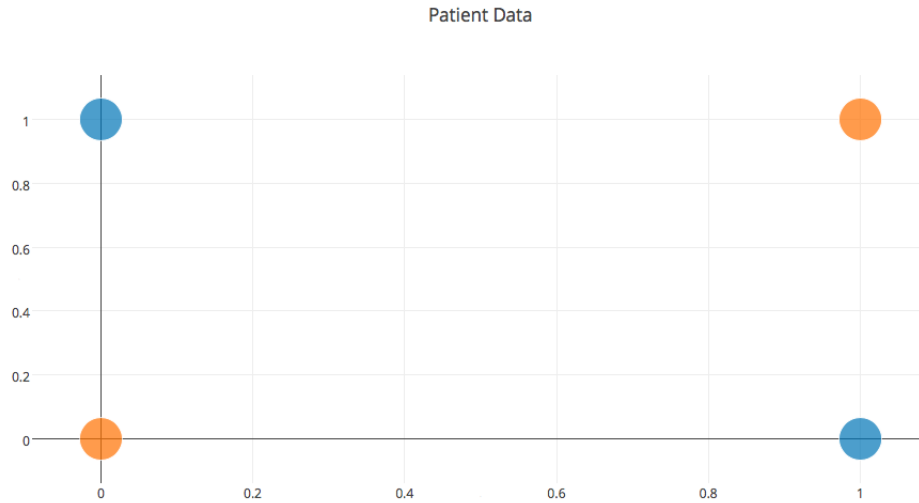


Figure 1: Patient Data

Answer to Question 1(c)

From below figures it is found that only $\psi = ((\phi_1 \text{ xor } \phi_2), \phi_2, 1)$ is linear separable. The associated $w = (-1, 0, -0.5)$

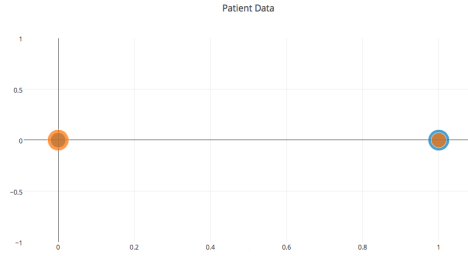


Figure 2: $\psi = (\phi_1^2, \phi_1\phi_2, 1)$

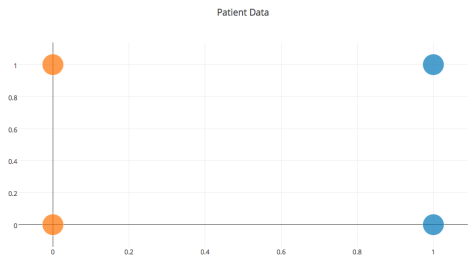


Figure 3: $\psi = ((\phi_1 \text{ xor } \phi_2), \phi_2, 1)$

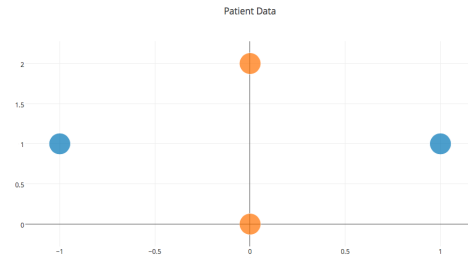


Figure 4: $\psi = (\phi_1 - \phi_2, \phi_1 + \phi_2, 1)$

Answer to Question 2(a)

$$\nabla_b f(w, b) = C \sum_{i=1}^n \frac{\partial L(x_i, y_i)}{\partial b}$$

Where:

$$\frac{\partial L(x_i, y_i)}{\partial b} = \begin{cases} 0 & \text{if } y_i(x_i w + b) \geq 1 \\ -y_i & \text{if } \textit{otherwise}. \end{cases}$$

Answer to Question 2(b)

Answer to Question 3(a)

Answer to Question 3(b)

Answer to Question 4(a)

Answer to Question 4(b)

Cover Sheet

Assignment Submission Fill in and include this cover sheet with each of your assignments. Assignments are due at 11:59pm. All students (SCPD and non-SCPD) must submit their homeworks via GradeScope (<http://www.gradescope.com>). Students can typeset or scan their homeworks. Make sure that you answer each question on a separate page. Students also need to upload their code at <http://snap.stanford.edu/submit>. Put all the code for a single question into a single file and upload it. Please do not put any code in your GradeScope submissions.

Late Day Policy Each student will have a total of *two* free late periods. *One late period expires at the start of each class.* (Homeworks are usually due on Thursdays, which means the first late periods expires on the following Tuesday.) Once these late periods are exhausted, any assignments turned in late will be penalized 50% per late period. However, no assignment will be accepted more than *one* late period after its due date.

Honor Code We strongly encourage students to form study groups. Students may discuss and work on homework problems in groups. However, each student must write down their solutions independently i.e., each student must understand the solution well enough in order to reconstruct it by him/herself. Students should clearly mention the names of all the other students who were part of their discussion group. Using code or solutions obtained from the web (github/google/previous year solutions etc.) is considered an honor code violation. We check all the submissions for plagiarism. We take the honor code very seriously and expect students to do the same.

Your name: _____

Email: _____ **SUID:** _____

Discussion Group (People with whom you discussed ideas used in your answers):

On-line or hardcopy documents used as part of your answers:

I acknowledge and accept the Honor Code.

(Signed)_____