

This assignment contains 7 questions worth a total of 22 points. Because assignments will help you learn things that are complementary to what we covered in class, they should be completed on your own. Otherwise, you will not learn from taking this course and you are harming yourself.

**Problem 1** Let  $(w, b) \in \mathbb{R}^d \times \mathbb{R}$  and  $x \in \mathbb{R}^d$ . Assume that  $\|w\|_2 = 1$ . Let us consider the point  $v$  defined by  $v = x - (w \cdot x + b)w$ .

1. (2 points) Show that  $w \cdot v + b = 0$
2. (3 points) Using the previous question, prove that the following holds:

$$\min_{u \in \mathbb{R}^d} \text{ s.t. } w \cdot u + b = 0 \quad \|x - u\|_2 \leq |w \cdot x + b|$$

3. (4 points) Let  $u \in \mathbb{R}^d$  such that  $w \cdot u + b = 0$ . Show that  $\|x - u\|_2 \geq \|x - v\|_2$ .
4. (2 points) Use the above results to find the analytical expression for the  $\ell_2$  distance between a point  $x$  and the hyperplane defined by  $w \cdot u + b = 0$  for all  $u \in \mathbb{R}^d$ .

**Problem 2** Recall the perceptron algorithm from class. It can be found in lecture notes posted on the course website.

1. (2 points) Let us modify the update rule to add a learning rate  $\alpha$  which multiplies the update applied to weights. The update rule becomes:  $w' \leftarrow w + \alpha t^{(i)} x^{(i)}$ . Show that the value of the learning rate does not change the perceptron's prediction after an update.
2. (8 points) Implement this modified perceptron algorithm in Python and NumPy. Include a copy of your code in your submission for this assignment.
3. (1 point) Report the test error of your perceptron on the breast cancer dataset included with sklearn. Split the dataset (which includes 569 points) in 450 training examples and 119 test examples. One way to do that is as follows:

```
from sklearn.datasets import load_breast_cancer
breast_cancer = load_breast_cancer()
X = breast_cancer.data
Y = breast_cancer.target
train_X = X[:450]
test_X = X[450:]
train_Y = Y[:450]
test_Y = Y[450:]
```

Note that TAs will not have time to help with setting up your machines. If you are unable to install Python, NumPy, or sklearn on your machine, you can use the notebooks provided for free on Colab: <https://colab.research.google.com>.