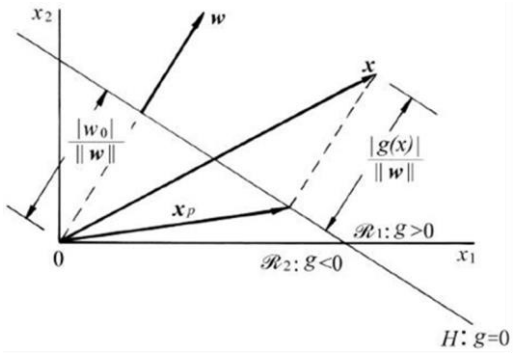
Problem Set 1

Sept. 16, 2021

1. Investigate the technologies that you may meet or use in everyday life, find among them an example of a machine learning technology. Write a short essay (about half-page) to introduce it and argue why it is a machine learning technology.

2. Write down the mathematical equation of a linear decision boundary in *d*-dimensional space $\boldsymbol{x} \in R^d$, derive the distance from any point $\boldsymbol{x}$ to the decision boundary, and the distance from the original point to the decision boundary.



3. Suppose there are three products for testing the infection of the SARS-Cov-2 virus. Product-A has sensitivity of 70% and specificity of 100%. Product-B has sensitivity 90% and specificity of 99%. Product-C has sensitivity of 99% and specificity of 70%. Suppose the three products are of the same cost. In a city of 1 million citizens, currently there have been ~100 confirmed positive cases. They are all sporadic cases without known epidemiological associations. Discuss the following questions:
   (1) As a citizen of the city, I'm suspecting that I might have had contact with someone who might has the virus. I'm thinking to take a test. Which product I should choose? I hope I can be treated timely if I really have got the virus, but I also do not want to make all my family and friends to be mistakenly quarantined because of a false alarm.
   (2) If I'm the public health officer of the city, and I plan to test all citizens, with the hope to take necessary measures to prevent the widespread of the pandemic in my city. But I also keep the minimum interruption on the normal social activities. What are the pros and cons of each product?

Due date: Sept. 22 (Wed.) 23:00 Beijing time