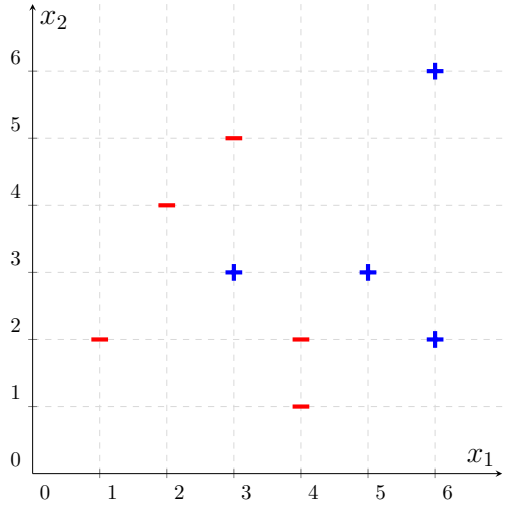
6.390 Introduction to Machine Learning

Recitation Week #4

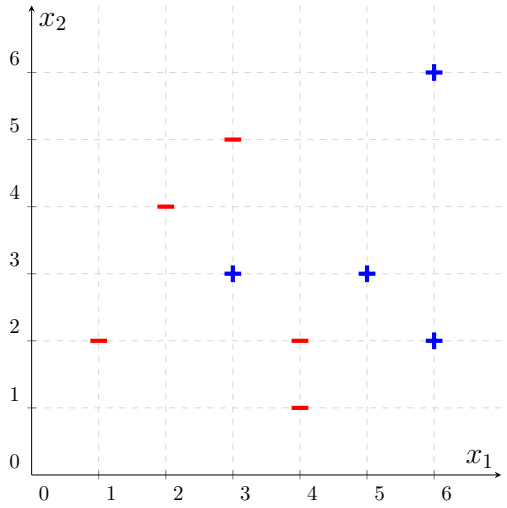
Issued September 26, 2022

1. Consider the training data set plotted below:



(a) Write down the formula from the notes for classification accuracy (hint: $0-1$ loss should appear in your formula).

(b) Draw a hyperplane that obtains the smallest training error in terms of highest classification accuracy. Be sure to also draw the normal vector.



Plot duplicated for convenience.

(c) Suppose we remove datapoints $(3, 3, +1)$ and $(4, 2, -1)$. Let two hypotheses be considered different if there exists a test point (i.e., not necessarily from the data set shown) that they would classify differently. How many different hypotheses are there that obtain zero training error? Explain your answer.

(d) Is there a reason to prefer some of these zero training error hypotheses over others? How might we find one or more of those?

2. For each of the following, determine if the statement is true or false and justify your reasoning.

   (a) If we take any linearly separable data set and add a new feature, it is still guaranteed to be linearly separable.

   ◯ True ◯ False

   (b) If we take any linearly separable data set and remove a feature, it is still guaranteed to be linearly separable.

   ◯ True ◯ False

   (c) If we take any data set that is not linearly separable and remove a feature, it is still guaranteed to not be linearly separable.

   ◯ True ◯ False

   (d) If we take any data set that is not linearly separable and remove a data point, it is still guaranteed to not be linearly separable.

   ◯ True ◯ False

3. Beatriz used logistic regression on a data set derived from people living in Massachusetts to learn a linear logistic classifier $\sigma(\theta^\top x + \theta_0)$ giving the probability that an adult with features x will develop heart disease in the next decade. Let θ^* and θ_0^* be the set of parameters that Beatriz learns.

   Her friend, John, would like to use the same logistic regression classifier (i.e., with the θ^* and θ_0^* learned by Beatriz) to make predictions for people living in Norway. However, he notices that heart disease is much less common in Norway and thinks that the model may need to be adjusted to account for this. For now, he decides to follow Beatriz in using logistic regression and decides to consider only changes to the offset θ_0 (that is, he always uses the same $\theta = \theta^*$ as Beatriz).

   (a) Consider a specific patient with feature vector x. How could John adjust the offset parameter, relative to the θ_0^* learned by Beatriz, so as to make smaller the probability of this patient developing heart disease?

   (b) John realizes that choosing the right value of θ_0 is tricky since he doesn't have access to any labeled data from Norway. John tells Beatriz that he only plans to use the model to find the 10% of individuals with highest probability of developing heart disease so that he can closely follow them and make sure they are tested appropriately. "Aha!", says Beatriz. "In that case, any value of θ_0 would suffice, and you can simply make use of my original linear logistic classifier!" Is Beatriz correct? Why or why not?