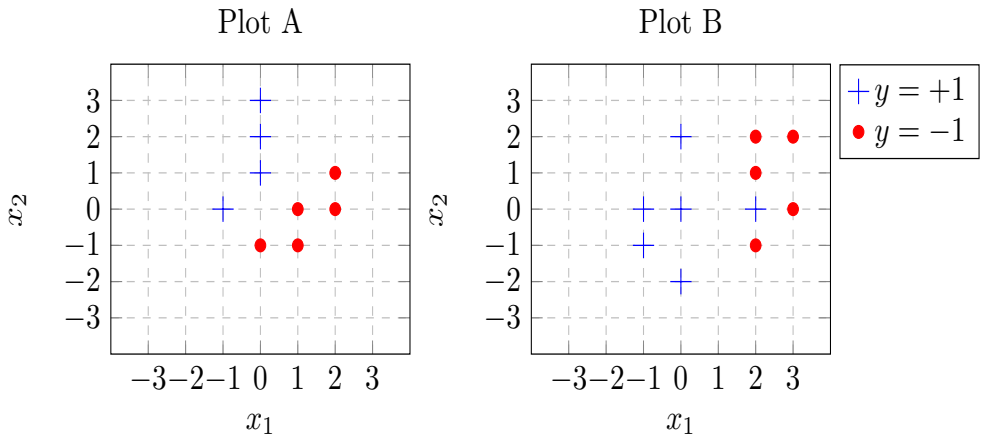
Name: _____

Linear Classifiers

1. (14 points) In the plots below, we give you 2D points with $+1$ and -1 labels.



Answer the following questions for both plot A and plot B:

(a) Using a linear separator $h(p; \theta, \theta_0) = sign(\theta^\top p + \theta_0)$, what is the minimum possible number of misclassified points?

Plot A:

Solution: 0. A separator which achieves this is given in the solution to part (b).

Plot B:

Solution: 1. A separator which achieves this is given in the solution to part (b).

(b) What are the values of $\theta \in \mathbb{R}^2$ and $\theta_0 \in \mathbb{R}$ that define your separator?

Plot A:

Solution: $\theta = [-1, 1]^T$, $\theta_0 = 0$, and $h(p) = +1$.

Plot B:

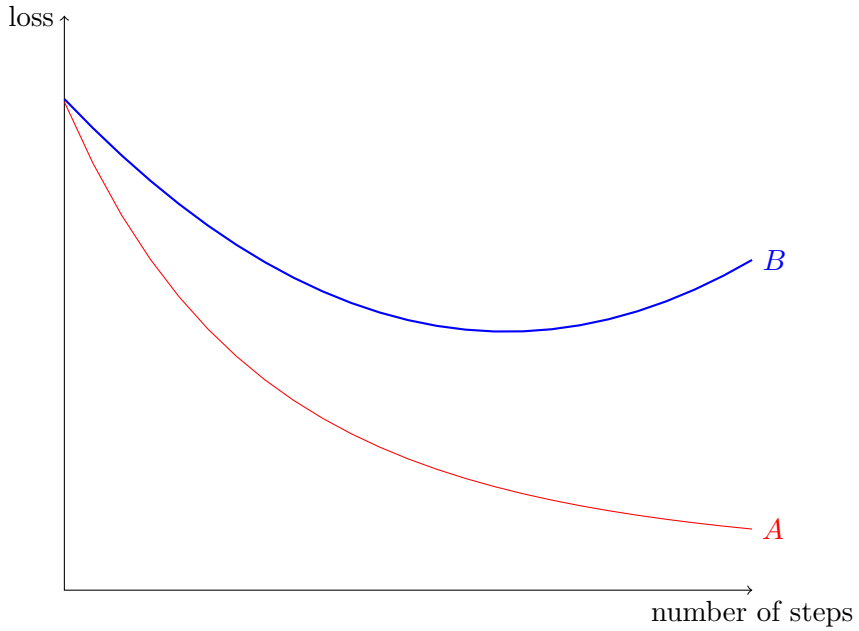
Solution: $\theta = [-1, 0]^T$, $\theta_0 = 1$, and $h(p) = -1$.

(c) For a given point p, what does $\frac{\theta^\top p + \theta_0}{\|\theta\|}$ intuitively represent?

Solution: Signed distance from the point p to the separating line.

Name: _____

Consider the following plot from the previous classification task. The two curves show the train and test error vs. the number of steps in the optimization algorithm.



(d) Assign the appropriate labels:
Test error (select one): ◯ A √ **B**
Train error (select one): √ **A** ◯ B

> **Solution:** The training error will always be decreasing (given that our step size is small enough) as our optimization algorithm is directly minimizing it. Our test error will start increasing after a certain point due to overfitting.

(e) Which of the following options can improve the final performance of the trained classifier on the test data set? Note: augmentation of a data set refers to taking the existing data set and adding many points which are slightly perturbed versions of the original points. Select all that apply.

√ **A. Augment the training data set and retrain the classifier.**

◯ B. Augment the test data set and retrain the classifier.

√ **C. Terminate the training process earlier.**

√ **D. Add a penalty on the magnitude of the parameter values and retrain the classifier.**

Solution: Augmenting the training data set can help improve generalization, as the requirement of correctly classifying the noisy, new points tends to keep the new classifier boundary away from the original points. As the margin tends to increase in size, generalization improves.

Augmenting the test set does not change the classifier that was trained– it only gives you a less accurate idea of what the generalization error will be.

Terminating training early can prevent overfitting (e.g. stopping near the minimum of the curve B would have been ideal in order to minimize test error).

Regularizing our parameters tends to prevent overfitting, as well.