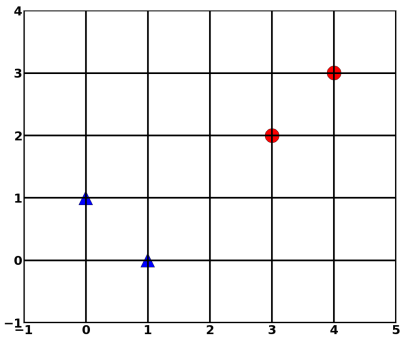
1. Support Vector Machines

   a. We typically frame an SVM problem as trying to *maximize* the margin. Explain intuitively why a bigger margin will result in a model that will generalize better, or perform better in practice.

   b. Show that the width of an SVM slab with linearly separable data is $\frac{2}{\|w\|}$.

   c. You're presented with the following set of data (triangle = +1, circle = -1):



   Find the equation (by hand) of the hyperplane $\vec{w}^T x + b = 0$ that would be used by an SVM classifier. Which points are support vectors?

2. What's the difference between the perceptron algorithm and the hard-margin SVM algorithm?

3. What's the difference between the hard-margin and the soft-margin SVM? How does the hyperparameter C affect the solution to the soft-margin SVM?

4. Matrix Calculus (Linear Regression)

   Let $X \in \mathbb{R}^{n \times d}$ be a data matrix and $y \in \mathbb{R}^n$ be the corresponding vector of labels. What is the weight vector $\theta \in \mathbb{R}^d$ that minimizes the quadratic loss between the predicted labels $X\theta$ and the actual labels y?