1 Fun with Lagrange Multipliers

(a) Minimize the function

$$f(x,y) = x + 2y$$

such that

$$x^2 + y^2 = 3.$$

(b) Minimize the function

$$f(x,y,z) = x^2 - y^2$$

such that

$$x^2 + 2y^2 + 3z^2 = 1.$$

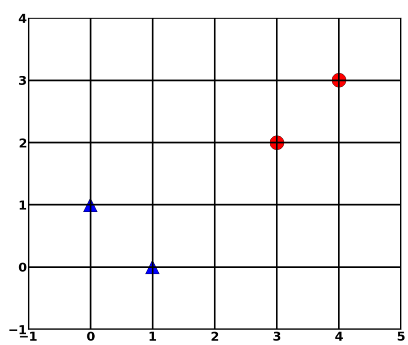
2 Support Vector Machines

(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) Will moving points which are not support vectors further away from the decision boundary effect the SVM's hinge loss?

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

(d) You are 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?

3 Simple SGD updates

Let us consider a simple least squares problem, where we are interested in optimizing the function

$$F(w) = \frac{1}{2n}\|Aw - y\|_2^2 = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}(a_i^\top w - y_i)^2.$$

(a) What is the closed form OLS solution? **What is the time complexity of computing this solution in terms of flops?**

(b) Write down the gradient descent update. **What is the time complexity of computing an ε optimal solution?**

(c) Write down the stochastic gradient descent update. **What is the time complexity of computing an ε optimal solution?** You may want to quickly go through a derivation here. **What happens when $Aw^* = y$?**

Discuss why you would use any of these methods for your problem.

(d) Write down the SGD update for logistic regression on two classes

$$F(w) = \frac{1}{n}\sum_{i=1}^{n} y_i \log\frac{1}{\sigma(w^\top x_i)} + (1 - y_i)\log\frac{1}{1 - \sigma(w^\top x_i)}.$$

Discuss why this is equivalent to minimizing a "cross-entropy" loss.