1. A ball in R^d of radius r centered at point $x_0 \in R^d$ is given by a set
$$B_R(x_0) = \{x : \|x - x_0\|_2^2 \leq r^2\}.$$
Formulate an optimization problem of finding the ball of minimal radius that contains a given set
$$S = \{x_1, x_2, \ldots, x_n\}.$$
Convert the problem to a dual form and show that solution can be obtained as a linear combination of a set S. Download data set `sphere.txt` and find a minimal enclosing sphere for this set. Give coefficient of the linear combination of points in S that define the sphere.

   If you are using `R`, you may want to consider package `quadprog` to solve the optimization problem.

2. We've seen that 2-norm soft margin SVM corresponds to the following optimization problem.
$$\begin{aligned} \min_{\xi,w,b} \quad & \langle w, w, \rangle + C\sum_{i=1}^{n} \xi_i^2 \\ \text{s.t.} \quad & y_i(\langle w, x_i\rangle + b) \geq 1 - \xi_i, i = 1, \ldots, n. \end{aligned}$$
A 1-norm soft margin SVM can be obtained by replacing the $\sum_{i=1}^{n} \xi_i^2$ term in the above problem with $\sum_{i=1}^{n} \xi_i$ and introducing appropriate constraints on ξ.

   (a) Formulate the appropriate optimization problem for 1-norm SVM.

   (b) What is the Lagrangian for this problem?

   (c) Write the dual optimization problem for 1-norm SVM.

   (d) Once $\alpha^\star$ is computed for the dual, how do we find $b^\star$?

   (e) What is the norm of $w^\star$?

3. Investigate the performance of SVM classifier in a binary classification problem. There are several implementations of SVM available. For example, if you are using `R`, you may consider using function `svm` in package `e1071`. More options can be found at `http://www.kernel-machines.org/software.html`. Consider two datasets:

   - Simulate a two-class classification problem with X uniformly distributed in $[-4, 4]^2$ and $\mathbf{P}(Y = 1|X = x) = 0.1 + 0.8I_{\{w^T x \geq 0\}}$.
   - Simulate a two-class classification problem with X uniformly distributed in $[-4, 4]^2$ and $\mathbf{P}(Y = 1|X = x) = 0.1 + 0.8I_{\{\cos(\|x\|_2^2) \geq 0\}}$.

   For each of the distributions above investigate the effect of the kernel and parameters (for example, of the parameter σ in the Gaussian kernel $k(x, z) = \exp(- \parallel x - z \parallel^2 / \sigma^2)$ or parameters d and r in polynomial kernel $k(x, z) = (\langle x, z\rangle + r)^d$, as well as the regularization constant C in standard formulation) on the decision boundary, number of support vectors and accuracy of classification. Visualize the results.

4. Download any publicly available dataset that interests you. You may look, for example, at UCI repository at `http://archive.ics.uci.edu/ml/`. Use cross validation to find optimal parameters. Use a hold-out set to estimate misclassification probability. Compare performance of the SVM classifier to a logistic regression (use function `glm` in `R`).

5. Use S&P data in `sp500x1.csv` file available on the course web page to train an SVM classifier. Description of the data is given in file `sp500x1description.txt`.

   (a) Find optimal parameter values for the classifier and estimate its performance.

   (b) (optional) Use WRDS datasets to extract additional predictors and try to improve the prediction accuracy.