Exercise 1 (Quadratics And Least Squares) Consider the two dimensional quadratic function, \( f : \mathbb{R}^2 \to \mathbb{R} \) given by:
\[
f(w) = w^\top A w - 2b^\top w + c
\]
where \( A \in \mathbb{S}^2_+ \), \( b \in \mathbb{R}^2 \) and \( c \in \mathbb{R} \).

1. Explain why the function \( f \) is convex.

2. Assume \( c = 0 \). Give a concrete example of a matrix \( A > 0 \) and a vector \( b \) such that the point \( w^* = [-1 \ 1]^\top \) is the unique minimizer of the quadratic function \( f(w) \).

3. Assume \( c = 0 \). Give a concrete example of a matrix \( A \succeq 0 \), and a vector \( b \) such that the quadratic function \( f(w) \) has infinitely many minimizers and all of them lie on the line \( w_1 + w_2 = 0 \).

4. Assume \( c = 0 \). Give a concrete example of a non-zero matrix \( A \succeq 0 \) and a vector \( b \) such that the quadratic function \( f(w) \) tends to \(-\infty\) as we follow the direction defined by the vector \([1 \ 0]^\top\).

5. Say that we have the data set \( \{(x^{(i)}, y^{(i)})\}_{i=1,...,n} \) of features \( x^{(i)} \in \mathbb{R}^2 \) and values \( y^{(i)} \in \mathbb{R} \). Define \( X = [x^{(1)} \ldots x^{(n)}]^\top \) and \( y = [y^{(1)} \ldots y^{(n)}]^\top \). In terms of \( X \) and \( y \), find a matrix \( A \), a vector \( b \) and a scalar \( c \), so that we can express the sum of the square losses \( \sum_{i=1}^{n} (w^\top x^{(i)} - y^{(i)})^2 \) as the quadratic function \( f(w) = w^\top A w - 2b^\top w + c \).

6. Which of the following can be true for the minimization of the sum of the square losses of part (5):
   
   (a) It can have a unique minimizer.
   
   (b) It can have infinitely many minimizers, all of them lying on a single line.
   
   (c) It can be unbounded from below, i.e. there is some direction so that if we follow this direction the loss tends asymptotically to \(-\infty\).

Exercise 2 (Solving Least Squares with CVX) 1. Use the standard normal distribution in order to generate a random \( 16 \times 8 \) matrix \( X \), and a random \( 16 \times 1 \) vector \( y \). Then use CVX in order to solve the least squares problem:
\[
\min_{w \in \mathbb{R}^8} \left\| Xw - y \right\|_2^2
\]
Check your answer by comparing with the analytic least squares solution.
2. Now assume that we are interested in finding a binary valued vector \(w\) for the least squares problem, i.e. we would like to solve

\[
p^* = \min_{w \in \mathbb{R}^8} \|Xw - y\|_2^2 : w_i \in \{0, 1\}, i = 1, \ldots, 8
\]

Note that this problem is not convex, but we can form the following convex relaxation

\[
p^*_{\text{int}} = \min_{w \in \mathbb{R}^8} \|Xw - y\|_2^2 : 0 \leq w_i \leq 1, i = 1, \ldots, 8
\]

Use CVX to find \(p^*_{\text{int}}\). What is the relation between \(p^*\) and \(p^*_{\text{int}}\)?

3. Finally use CVX to solve the LASSO problem

\[
\min_{w \in \mathbb{R}^8} \|Xw - y\|_2^2 + \lambda \|w\|_1
\]

where \(\lambda > 0\) is a hyper-parameter. Use values of \(\lambda\) in the interval \([10^{-4}, 10^6]\), and create a plot of each coordinate \(w_i\) of the optimal vector \(w\) versus the corresponding hyper-parameter \(\lambda\).

Exercise 3 (A Simple Case Of LASSO) Say that we have the data set \(\{(x^{(i)}, y^{(i)})\}_{i=1,\ldots,n}\) of features \(x^{(i)} \in \mathbb{R}^d\) and values \(y^{(i)} \in \mathbb{R}\). Define \(X = [x^{(1)} \ldots x^{(n)}]^{\top}\) and \(y = [y^{(1)} \ldots y^{(n)}]^{\top}\). For the sake of simplicity, assume that the data has been centered and whitened so that each feature has mean 0 and variance 1 and the features are uncorrelated, i.e. \(X^{\top}X = nI\).

Consider the linear least squares regression with regularization in the \(\ell_1\)-norm, also known as LASSO:

\[
w^* = \arg \min_{w \in \mathbb{R}^d} \|Xw - y\|_2^2 + \lambda \|w\|_1
\]

1. Decompose this optimization problem in \(d\) univariate optimization problems.

2. If \(w_i^* > 0\), then what is the value of \(w_i^*\)?

3. If \(w_i^* < 0\), then what is the value of \(w_i^*\)?

4. What is the condition for \(w_i^*\) to be 0?

5. Now consider the case of ridge regression, which uses the the \(\ell_2\) regularization \(\lambda \|w\|_2^2\).

\[
w^* = \arg \min_{w \in \mathbb{R}^d} \|Xw - y\|_2^2 + \lambda \|w\|_2^2
\]

What is the new condition for \(w_i^*\) to be 0? How does this differ from the condition obtained in part (4)? What does this suggest about LASSO?