CS/ECE/ME 532

Unit 5 Practice Problems

1. **Support vector machines**. We are trying to predict whether a certain chemical reaction will take place as a function of our experimental conditions: temperature, pressure, concentration of catalyst, and several other factors. For each experiment $i = 1, \dots, m$ we record the experimental conditions in the vector $a_i \in \mathbb{R}^n$ and the outcome in the scalar $b_i \in \{-1, 1\}$ ($+1$ if the reaction occurred and -1 if it did not). We will use an SVM to train our linear classifier. Namely, we solve:

$$\underset{x}{\text{minimize}} \quad \sum_{i=1}^{m} (1 - b_i a_i^\mathsf{T} x)_+ \qquad \text{where } (u)_+ = \max(0, u) \text{ is the soft thresholding operator}$$

a) Derive a gradient descent method for solving this problem. Explicitly give the computations required at each step. ***Note:*** you may ignore points where the function is non-differentiable.

SOLUTION: Using the definition of soft-threshold, we have:

$$(1 - b_i a_i^\mathsf{T} x)_+ = \begin{cases} 0 & \text{if } b_i a_i^\mathsf{T} x > 1 \\ 1 - b_i a_i^\mathsf{T} x & \text{if } b_i a_i^\mathsf{T} x < 1 \end{cases}$$

Therefore the gradient is given by:

$$\nabla_x (1 - b_i a_i^\mathsf{T} x)_+ = \begin{cases} 0 & \text{if } b_i a_i^\mathsf{T} x > 1 \\ -b_i a_i & \text{if } b_i a_i^\mathsf{T} x < 1 \end{cases}$$

We can write this compactly as $\nabla_x (1 - b_i a_i^\mathsf{T} x)_+ = -\frac{1}{2} b_i \left(1 + \text{sign}(1 - b_i a_i^\mathsf{T} x)\right) a_i$. A gradient descent algorithm involves the entire gradient and would look like:

1. initialize x_0
2. compute $x_{k+1} = x_k + \frac{\gamma}{2} \sum_{i=1}^{m} b_i \left(1 + \text{sign}(1 - b_i a_i^\mathsf{T} x)\right) a_i$ for $k = 0, 1, \dots$

b) Explain what happens to the algorithm if you land at an x_k that classifies all the points perfectly, and by a substantial margin.

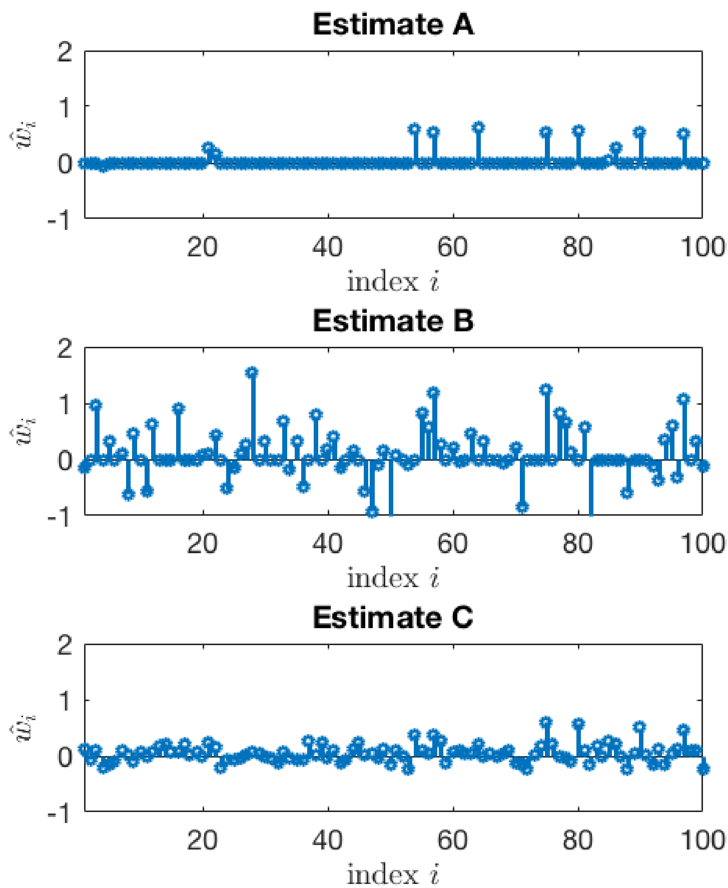
SOLUTION: If classification is perfect, this means $b_i a_i^\mathsf{T} x > 0$ for all i. If the margin is large enough so that $b_i a_i^\mathsf{T} x > 1$ as well, then the gradient will be zero. So the gradient descent iterations stop.

2. Regularization

a) We observe $n = 60$ training samples of the form $(\boldsymbol{x}_i, y_i) \in \mathbb{R}^p \times \mathbb{R}$ for $i = 1, \ldots, n$, where $p = 100$. For these samples, we compute the least-squares estimator, the ridge regression estimator with squared error loss, and the LASSO estimator. Write expressions for each of the three estimators (e.g., $\widehat{\boldsymbol{w}} = \arg\min \ldots$). **SOLUTION:**

$$\begin{aligned}
(\text{least squares})\widehat{w} &= \arg\min \|y - Xw\|_2^2 \\
(\text{ridge})\widehat{w} &= \arg\min \|y - Xw\|_2^2 + \lambda\|w\|_2^2 \\
(\text{lasso})\widehat{w} &= \arg\min \|y - Xw\|_2^2 + \lambda\|w\|_1
\end{aligned}$$

b) The three estimates are in the plots below. Identify which estimator is in which plot and explain your reasoning.



SOLUTION: A is Lasso – sparsest
B is least squares because norm of w is clearly larger than in C. C is ridge.

3. You use gradient-descent based iterative algorithms for finding the 2-by-1 vector of weights $\boldsymbol{w}$ that solves three different least-squares problems:

$$\text{A: } \min_{\boldsymbol{w}} ||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2$$

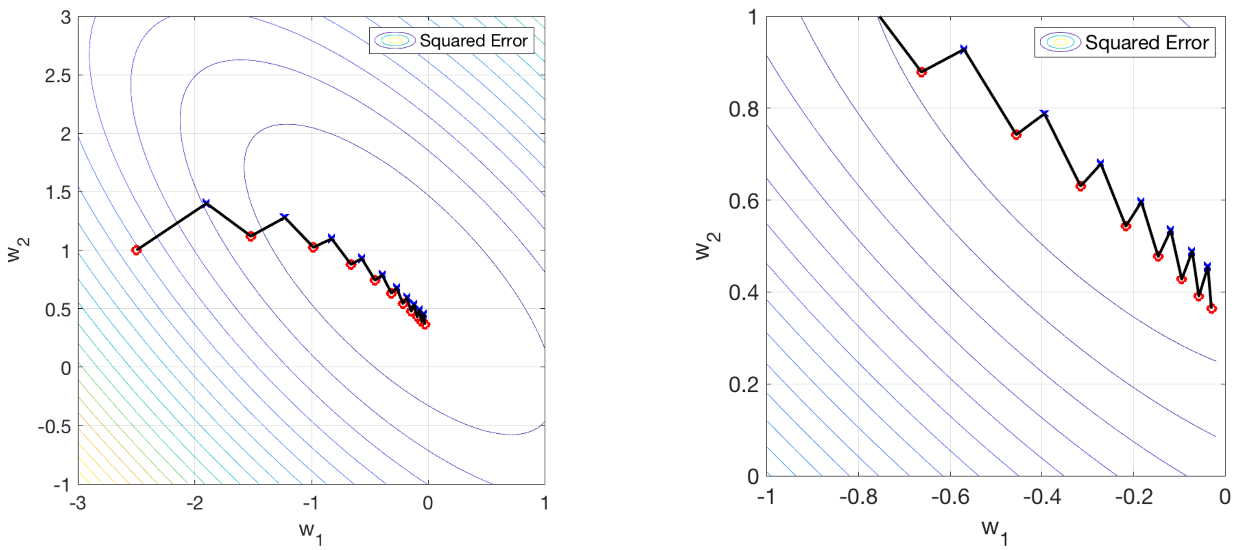
$$\text{B: } \min_{\boldsymbol{w}} \left\{||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2 + \lambda||\boldsymbol{w}||_2^2\right\}$$

$$\text{C: } \min_{\boldsymbol{w}} \left\{||\boldsymbol{X}\boldsymbol{w} - \boldsymbol{y}||_2^2 + \lambda||\boldsymbol{w}||_1\right\}$$

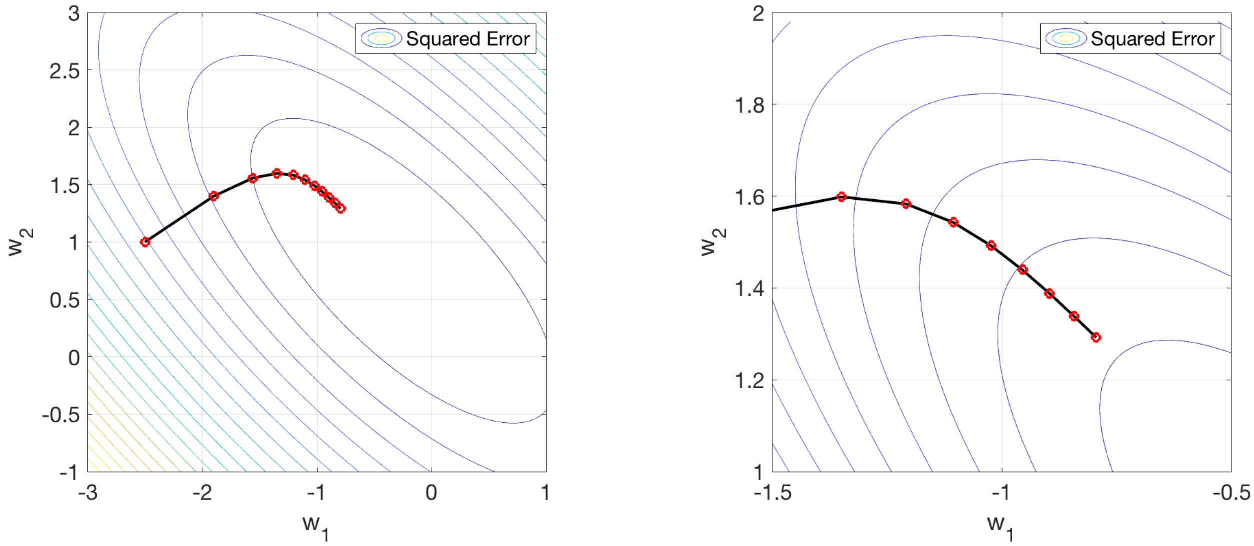
The graphs shown in each part below depict the trajectories of the weights for the first ten iterations. Identify which of these problems correspond to each of the trajectories. **Give a the reason for each of your answers.**

Note that $\boldsymbol{X}$ and $\boldsymbol{y}$ are the same in all cases. Successive iterations of the weights are denoted by the circles. The `x` symbol denotes the intermediate step within each iteration. The trajectory shown in the right panel is a closer view of a section of the trajectory shown in the left panel.

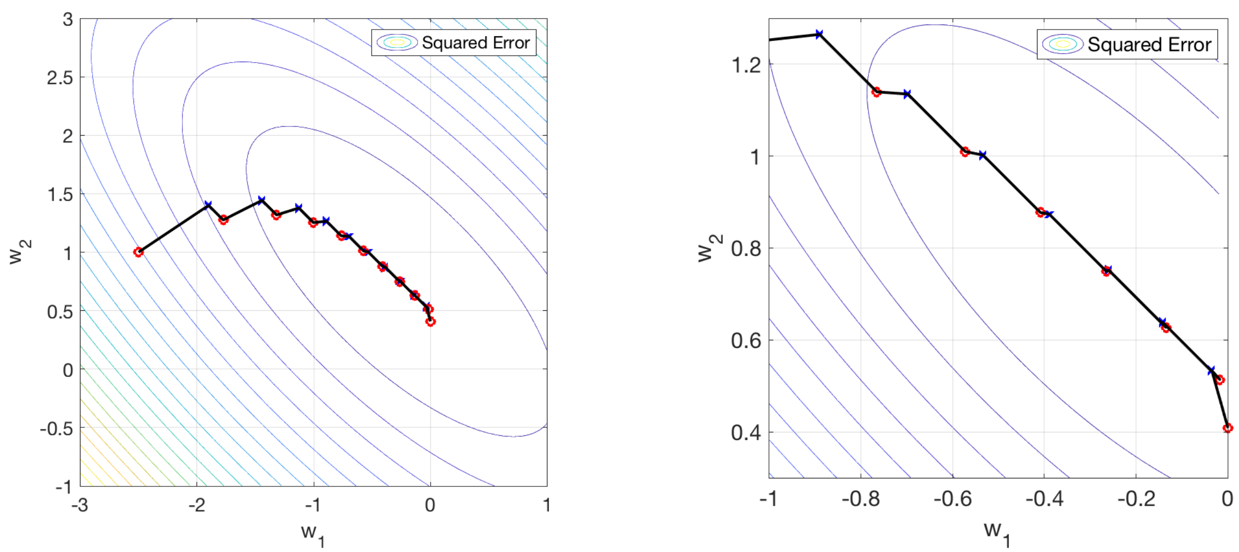
a) Answer: _________



b) Answer: _________



c) Answer: _________



SOLUTION:

a) B, Ridge Regression

b) A, Least Squares

c) C, LASSO