Georgia Institute of Technology

ECE 6254 - Assignment 9

Fall 2020 - v1.0

- There are 2 problems over 2 pages (including the cover page.
- The problems are not necessarily in order of difficulty.
- All problems are assigned the same weight in the overall grade. The questions in problems with many questions are worth less than the questions in problem with few questions.
- Each question is graded as follows: no credit without meaningful work, half credit for partial work, full credit if essentially correct.
- Unless otherwise specified, you should concisely indicate your reasoning and show all relevant work.
- The grade on each question is based on our judgment of your level of understanding as reflected by what you have written. If we cannot read it, we cannot grade it.
- Please use a pen and not a pencil if you handwrite your solution.
- **You must submit your assignment on Gradescope**.

Problem 1: Elastic-net regularizer

In class we discussed ridge regression and the LASSO. Another form of regularized least squares regression involves the so-called *elastic-net regularizer*, which corresponds to the optimization problem

$$\min_{\boldsymbol{\theta}} \|\mathbf{y} - \mathbf{X}\theta\|_2^2 + \lambda \left(\alpha \|\boldsymbol{\theta}\|_2^2 + (1-\alpha)\|\boldsymbol{\theta}\|_1 \right) \tag{1}$$

where both λ and α are scalar parameters set by the user. The elastic-net regularizer can be viewed as a compromise between the ℓ_2 and ℓ_1 penalties, being prone to both selecting variables (like the LASSO) and shrinking together the coefficients of correlated predictors (like ridge regression). Show that the elastic-net optimization problem is as a LASSO optimization problem with an augmented version of $\mathbf{y}$ and $\mathbf{X}$, which can be written as

$$\min_{\boldsymbol{\theta}} \|\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\boldsymbol{\theta}\|_2^2 + \tilde{\lambda}\|\boldsymbol{\theta}\|_1 \tag{2}$$

for some properly defined $\tilde{\mathbf{y}}$, $\tilde{\mathbf{X}}$, and $\tilde{\lambda}$ (*Hint:* $\tilde{\lambda}$ will be a combination of λ and α).

Problem 2: Specializing Tikhonov regularization

In this problem we consider the scenario seen in class, where x is drawn uniformly on $[-1, 1]$ and $y = \sin(\pi x)$, for which we are given $N = 2$ training samples. Here, we will consider an alternative approach to fitting a line to the data based on Tikhonov regularization. Specifically, we let

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \quad \mathbf{A} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix} \quad \boldsymbol{\theta} = \begin{bmatrix} b \\ a \end{bmatrix} \tag{3}$$

We will then consider Tikhonov regularized least squares estimators of the form

$$\hat{\boldsymbol{\theta}} \triangleq (\mathbf{A}^\intercal\mathbf{A} + \Gamma^\intercal\Gamma)^{-1}\mathbf{A}^\intercal\mathbf{y}. \tag{4}$$

[Q1] How should we set Γ to reduce this estimator to fitting a constant function (i.e., finding an $h(x)$ of the form $h(x) = b$)? (*Hint:* For the purposes of this problem, it is sufficient to set Γ in a way that just makes $a \approx 0$. To make $a = 0$ exactly requires setting Γ in a way that makes the matrix $\mathbf{A}^\intercal\mathbf{A} + \Gamma^\intercal\Gamma$ singular, but note that this does not mean that the regularized least-squares optimization problem cannot be solved; you must just use a different formula than the one in (4).

[Q2] How should we set Γ to reduce this estimator to fitting a line of the form $h(x) = ax + b$ that passes through the observed data points (x_1, y_1) and (x_2, y_2)?

[Q3] (Optional) Play around and see if you can find a (diagonal) matrix Γ that results in a smaller risk than either of the two approaches we discussed in class. You will need to do this numerically using Python or MATLAB. Report the Γ that gives you the best results. (You can restrict your search to diagonal Γ to simplify this.)