Homework 5: Optimization & Linear Regression

Instructor: Daniel L. Pimentel-Alarcón Due 03/28/2019

Problem 5.1. Compute the minimizer and minimum of

$$g(\beta) \;=\; 3\beta^2 - 5\beta + 4.$$

Problem 5.2. Let $f(\mathbf{W}, \boldsymbol{\beta})$ be a function with parameters $\mathbf{W}$ and $\boldsymbol{\beta}$. Suppose you want to find

$$\arg\min_{\mathbf{W},\boldsymbol{\beta}} \; f(\mathbf{W}, \boldsymbol{\beta}).$$

To this end you will use gradient descent, with initial points given by:

$$\mathbf{W}_0 = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, \qquad \boldsymbol{\beta}_0 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

If

$$\nabla \mathbf{W}\Big|_{t=0} = \begin{bmatrix} 0.2 & -0.3 \\ -0.1 & 0.2 \end{bmatrix}, \qquad \nabla \boldsymbol{\beta}\Big|_{t=0} = \begin{bmatrix} 0.1 \\ -0.2 \end{bmatrix},$$

what are the values of $\mathbf{W}_1$ and $\boldsymbol{\beta}_1$ after the first iteration of gradient descent?

Problem 5.3. In your preferred language, code gradient descent, and test it on g from Problem 5.1.

(a) Deliver your code.

(b) What value of η did you choose?

(c) What minimizer did you obtain?

(d) What minimum did you obtain?

(e) Do these values agree with your answer from Problem 5.1? How would you fix any discrepancies?

Problem 5.4. In the lecture notes we saw that the goal in linear regression is to find

$$\boldsymbol{\beta}^\star \;=\; \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^{\mathrm{D}+1}} \; \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2.$$

To this end we used matrix derivatives to conclude that

$$\boldsymbol{\beta}^\star \;=\; (\mathbf{X}^\mathsf{T}\mathbf{X})^{-1}\mathbf{X}^\mathsf{T}\mathbf{y}. \tag{5.1}$$

In this problem you will use scalar derivatives to show that this is correct. First observe that

$$\begin{aligned} g(\boldsymbol{\beta}) \;&=\; \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 \\ &=\; \sum_{\mathrm{i}=1}^{\mathrm{N}} (\mathrm{y_i} - \mathbf{x}_\mathrm{i}^\mathsf{T}\boldsymbol{\beta})^2 \\ &=\; \sum_{\mathrm{i}=1}^{\mathrm{N}} \left(\mathrm{y_i^2} - 2\mathrm{y_i}\mathbf{x}_\mathrm{i}^\mathsf{T}\boldsymbol{\beta} + (\mathbf{x}_\mathrm{i}^\mathsf{T}\boldsymbol{\beta})^2\right) \\ &=\; \sum_{\mathrm{i}=1}^{\mathrm{N}} \left[\mathrm{y_i^2} - 2\mathrm{y_i}\sum_{\mathrm{j}=1}^{\mathrm{D}} \beta_\mathrm{j}\mathrm{x_{ij}} + \left(\sum_{\mathrm{j}=1}^{\mathrm{D}} \beta_\mathrm{j}\mathrm{x_{ij}}\right)^2\right]. \end{aligned}$$

(a) Compute the derivative of $g(\boldsymbol{\beta})$ with respect to β_{k}.

(b) Set the derivative of $g(\boldsymbol{\beta})$ with respect to β_{k} to zero, and solve for β_{k}.

Problem 5.5. Consider the following vector $\mathbf{y}$, containing information about glucose level of three individuals, and the following data matrix $\mathbf{X}$ containing information about height and weight of the corresponding individuals:

$$\mathbf{y} = \begin{bmatrix} 110 \\ 140 \\ 180 \end{bmatrix}, \qquad \mathbf{X} = \begin{bmatrix} 180 & 150 \\ 150 & 175 \\ 170 & 165 \end{bmatrix}.$$

(a) Use (5.1) to find the vector $\boldsymbol{\beta}$ that best explains $\mathbf{y}$ as a linear function of $\mathbf{X}$.

(b) Use your answer from Part (b) in Problem 5.4 to find the vector $\boldsymbol{\beta}$ that best explains $\mathbf{y}$ as a linear function of $\mathbf{X}$.

(c) Do your answers from parts (a) and (b) coincide?

(d) Which option do you prefer?

(e) Given a new sample with feature vector $\mathbf{x} = [175 \;\; 170]^{\mathsf{T}}$, what would be your prediction for y?