IDNM 680: MATH HW 1

JAMES DELLA-GIUSTINA

Problem 1. Show that given Y and X are constants the loss function of linear regression is a convex function in terms of $\boldsymbol{\beta}$

$$L(\boldsymbol{\beta}) = (Y - X\boldsymbol{\beta})^2$$

where $Y = (y_1, y_2, \ldots, y_n)^\top$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)^\top$

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1k} \\ 1 & x_{21} & x_{22} & \ldots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nk} \end{bmatrix}$$

Solution 1. $L(\beta) = (Y - X\beta)^2$ must be written as

$$L(\beta) = (Y - X\beta)^\top (Y - X\beta) \equiv \left(-\beta^\top X^\top + Y^\top\right)(Y - X\beta)$$

Otherwise matrix multiplication would not be possible, i.e.

$$(Y - X\beta)(Y - X\beta) = (n \times 1)(n \times 1)$$
$$\Rightarrow\Leftarrow \text{ dimension mismatch}$$

So we have

$$L(\beta) = Y^\top Y - Y^\top X\beta - (X\beta)^\top Y + (Xp)^\top X\beta$$

Since

$$(X\beta)^\top Y = \beta^\top X^\top Y = \left(\beta^\top X^\top Y\right)^\top = Y^\top X\beta \wedge (X\beta)^\top X\beta = \beta^\top X^\top X\beta$$

then

$$\begin{aligned} &= \|Y\|^2 - Y^\top X\beta - \beta^\top X^\top Y + (X\beta)^\top X\beta \\ &= \|Y\|^2 - 2\beta^\top \left(X^\top Y\right) + \beta^\top X^\top X\beta \end{aligned}$$

Now, taking the gradient is the same as taking the derivative with respect to β, since X and Y are viewed as constants, so

$$\nabla L(\beta) = -2X^\top Y + X^\top X\beta := 0$$

$$X^\top X\beta = 2X^\top Y \Rightarrow \quad \therefore \beta = 2\left(X^\top X\right)^{-1} X^\top Y \qquad \text{is a local extremum of the loss function } L(\beta).$$

Now, find the Hessian by taking the gradient once more

$$\begin{aligned} H &= \nabla\left(-2X^\top Y + X^\top X\beta\right) \\ H &= X^\top X \\ &= \|X\|^2 \geq 0 \text{ for } \forall X \in \mathbb{M}_{n\times k+1}(\mathbb{R}) \end{aligned}$$

$$\therefore H \text{ is semipostive definite} \iff L(\beta) \text{ is convex.}$$

Date: Spring 2023.