| **INT 3405: Machine Learning** | **First Semester 2023-2024** |
|---|---|

Week name

*Student Name: Le Hoang Lan* — *UID: 22028013*

- Feel free to talk to other students in the class when doing the homework. You should, however, write down your solution yourself. You also must indicate on each homework with whom you collaborated and cite any other sources you use including Internet sites.
- You will write your solution in LaTeX and submit the pdf file in zip files, including relevant materials, through courses.uet.vnu.edu.vn
- Dont be late.

# 1 Homework 1 - 10pts

We have the loss function of Lasso Regression:

$$L(w) = \frac{1}{2n}\sum_{i=1}^{n}(\sum_{j=1}^{m} w_j x_{ij} + w_0 - y_i) + \lambda \sum_{j=1}^{m} |w_j| \tag{1}$$

or,

$$L(\mathbf{w}) = J(\mathbf{w}) + \lambda \left|\boldsymbol{w}\right| \tag{2}$$

We want to minimize the above loss function. However, $L(\mathbf{w})$ is not differientable at$\mathbf{w}$ that has any $w_j = 0$ due to the absolute values. Hence, Lasso Regression does not have any closed-form solution, or analytical solution. It is also not possible to directly apply Gradient Descent to Lasso Regression; we have to apply some variation of the method, for example, proximal gradient descent or subgradient method.

Subgradient method: The method is quite the same as the graident descent method. The only difference is that we will use the subgradient to update $\mathbf{w}$ instead of the gradient.

It can be seen that $L(\mathbf{w})$ is a convex function.

A subgradient of a convex function $f$ at $x$ is any vector $\mathbf{g} \in R^n$ such that

$$f(y) \geq f(x) + \mathbf{g}^T(y - x), \forall y \tag{3}$$

If $f$ is differentiable at $x$, then $\mathbf{g} = \bigtriangledown f(x)$.

The subgradient of $f(w) = |w|$ is:

$$\bigtriangledown^{sub} f(w) = \begin{cases} 1 & \text{if } w > 0 \\ w \in [-1, 1] & \text{if } w = 0 \\ -1 & \text{if } w < 0 \end{cases} \tag{4}$$

Since the subgradient of $|w|$ is familiar to the sign of $w$, we can use the sign of $w$ as the subgradient of $|w|$.

We obtain the subgradient of $L(w)$:

$$\bigtriangledown^{sub} L(\mathbf{w}) = \bigtriangledown J(\mathbf{w}) + \lambda sign(\mathbf{w}) \tag{5}$$

Here is the process of subgradient descent:

1. Initialize $\mathbf{w}$.

2. repeat {

   $\mathbf{w} := \mathbf{w} - \alpha(\bigtriangledown J(\mathbf{w}) + \lambda sign(\mathbf{w}))$

   } until convergence.

As $\lambda$ goes to infinity, the estimated weights $\mathbf{w}$ will go to 0, and the model will be come a horizontal line pararelling to $Ox$.

# 2 Homework 2 - 10pts

The inverse matrix of a matrix $A$ can be computed as follow:

$$A^{-1} = \frac{1}{|A|} Adj(A) \tag{6}$$

$|A|$ is the determinant of matrix $A$.

1. 2 data points and 2 features - 2x2 matrix For a 2x2 matrix, the inverse matrix is:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \tag{7}$$

2. 3 data points, 3 features - 3 x 3 matrix

   $Adj(A) = \begin{bmatrix} A_{22}A_{33} - A_{23}A_{32} & A_{13}A_{32} - A_{12}A_{33} & A_{23}A_{12} - A_{22}A_{13} \\ A_{31}A_{23} - A_{21}A_{33} & A_{11}A_{33} - A_{13}A_{31} & A_{21}A_{13} - A_{23}A_{11} \\ A_{21}A_{32} - A_{31}A_{22} & A_{11}A_{32} - A_{31}A_{12} & A_{11}A_{22} - A_{21}A_{12} \end{bmatrix}$

   $|A| = A_{11} * (A_{22}A_{33} - A_{23}A_{32}) + A_{12} * A_{13}A_{32} - A_{12}A_{33} + A_{13} * (A_{23}A_{12} - A_{22}A_{13})$

   Hence:

$$A^{-1} = \frac{1}{A_{11} * (A_{22}A_{33} - A_{23}A_{32}) + A_{12} * A_{13}A_{32} - A_{12}A_{33} + A_{13} * (A_{23}A_{12} - A_{22}A_{13})} Adj(A) \tag{8}$$

3. Data that has $n$ data points and $m$ features and $n \neq m$.

   It is impossible to find the inverse matrix in this case. We can only find the inverse matrix of a square matrix.

4. $n$ data points, $n$ features and $n > 3$

   We can apply the formula for inverse matrix. However, it may take a long time to compute.