HW1- Theory

* Vectors are denoted by **boldfaced** characters, matrices by **BOLDFACED CAPITAL** letters.

1. Prove Normal Equations:
   Given a training set $\mathcal{S} = \{\mathbf{X}, \mathbf{y}\}$, a linear hypothesis class $\{h_{\boldsymbol{\theta}}(\mathbf{x}) = \sum_{j=1}^{N} \theta_j x_j\}$ and the mean squared error loss function:

$$\mathcal{L} = \frac{1}{2M} \sum_{i=1}^{M} \left(h_{\boldsymbol{\theta}}(\mathbf{x_i}) - y_i \right)^2$$

   prove that $\boldsymbol{\theta}$ that minimizes $\mathcal{L}$ satisfies:

$$\mathbf{X}^T \mathbf{X} \boldsymbol{\theta} = \mathbf{X}^T \mathbf{y}$$

   where: $\quad \mathbf{x_i}, \boldsymbol{\theta} \in \mathbb{R}^N, \qquad \mathbf{y} \in \mathbb{R}^M, \qquad \mathbf{X} = \begin{bmatrix} — \ \mathbf{x_1}^T \ — \\ — \ \mathbf{x_2}^T \ — \\ \vdots \\ — \ \mathbf{x_M}^T \ — \end{bmatrix}, \qquad M \geq N$

2. Unique solution:
   Show that a unique solution for linear regression exists iff the features are not linearly dependent. Namely, show that a unique solution:

$$\underset{\boldsymbol{\theta}}{\operatorname{argmin}} \, \mathcal{L} = \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{y}$$

   exists iff $\mathbf{X}$ has full column rank.