DL4CV - Assignment 1

Itai Antebi
204817498

April 3, 2021

1. \

$$h_\theta\left(x_i\right)=\sum_{j=1}^{N}\theta_j x_{ij}$$

$$L=\frac{1}{2M}\sum_{i=1}^{M}\left(h_\theta\left(x_i\right)-y_i\right)^2=\frac{1}{2M}\sum_{i=1}^{M}\left(\sum_{j=1}^{N}\theta_j x_{ij}-y_i\right)^2$$

First Derivitive:

$$\begin{aligned}
\frac{\partial L}{\partial\theta_k}&=\frac{1}{2M}\sum_{i=1}^{M}2\left(\sum_{j=1}^{N}\theta_j x_{ij}-y_i\right)x_{ik}\\
&=\frac{1}{M}\sum_{i=1}^{M}\left(\sum_{j=1}^{N}\theta_j x_{ij}-y_i\right)x_{ik}\\
&=\frac{1}{M}\sum_{i=1}^{M}\left(\langle\theta,x_i\rangle-y_i\right)x_{ik}\\
&=\frac{1}{M}\left\langle\begin{pmatrix}\langle\theta,x_1\rangle-y_1\\ \vdots\\ \langle\theta,x_M\rangle-y_M\end{pmatrix},\begin{pmatrix}x_{1k}\\ \vdots\\ x_{Mk}\end{pmatrix}\right\rangle\\
&=\frac{1}{M}\left\langle X\theta-y,\left(X^T\right)_k\right\rangle\\
&=\frac{1}{M}\left\langle\left(X^T\right)_k,X\theta-y\right\rangle
\end{aligned}$$

The last equation is correct because the vectors $\left(X^T\right)_k$ and $X\theta-y$ are $\in R^M$ (and not $\in C^M$).

$$\nabla L=\frac{1}{M}X^T\left(X\theta-y\right)$$

Second Derivitive:

$$\begin{aligned}
\frac{\partial^2 L}{\partial\theta_{k_1}\partial\theta_{k_2}} &= \frac{\partial}{\partial\theta_{k_2}}\left(\frac{\partial L}{\partial\theta_{k_1}}\right) \\
&= \frac{\partial}{\partial\theta_{k_2}}\left(\frac{1}{M}\sum_{i=1}^{M}\left(\sum_{j=1}^{N}\theta_j x_{ij} - y_i\right)x_{ik_1}\right) \\
&= \frac{1}{M}\sum_{i=1}^{M}\frac{\partial}{\partial\theta_{k_2}}\left(\left(\sum_{j=1}^{N}\theta_j x_{ij} - y_i\right)x_{ik_1}\right) \\
&= \frac{1}{M}\sum_{i=1}^{M}\left(x_{ik_2}x_{ik_1}\right)
\end{aligned}$$

In other terms:

$$\nabla^2 L_{i,j} = \frac{\partial^2 L}{\partial\theta_i\partial\theta_j} = \frac{1}{M}\sum_{k=1}^{M}\left(x_{ki}x_{kj}\right) = \frac{1}{M}\left\langle \left(X^T\right)_i, \left(X^T\right)_j\right\rangle$$

$$\nabla^2 L = \frac{1}{M}X^T X$$

Convexity:

$$\begin{aligned}
\forall v \neq 0 \ \left\langle\left(\nabla^2 L\right)v, v\right\rangle &= \left\langle\left(\frac{1}{M}X^T X\right)v, v\right\rangle \\
&= \frac{1}{M}\left\langle\left(X^T X\right)v, v\right\rangle \\
&= \frac{1}{M}\left\langle X^T\left(Xv\right), v\right\rangle \\
&= \frac{1}{M}\left\langle Xv, Xv\right\rangle \\
&= \frac{1}{M}\left|Xv\right|^2 \geq 0
\end{aligned}$$

We have proven that $\nabla^2 L$ is positive semi-definite, and therefore L is a convex function of θ.

Main proof:

$$\begin{aligned}
\theta \text{ minimizes } L &\iff \nabla L\left(\theta\right) = 0 \\
&\iff \frac{1}{M}X^T\left(X\theta - y\right) = 0 \\
&\iff X^T\left(X\theta - y\right) = 0 \\
&\iff X^T X\theta = X^T\theta
\end{aligned}$$

The first $\iff$ stems from the fact that L is a convex function.

So, we have proven that θ minimizes $L \iff X^T X\theta = X^T\theta$.
Note that the question only required proofing the $\Rightarrow$ statement, but the converse direction is required for question 2.

2. **Sentence:** $A \in R^{M\times N} \Rightarrow rank\left(A\right) = rank\left(A^T A\right)$.
   Prove:

$$x \in N_A \Rightarrow Ax = 0 \Rightarrow A^T Ax = A^T 0 = 0 \Rightarrow x \in N_{A^T A}$$

$$N_A \subseteq N_{A^T A}$$

$$x \in N_{A^T A} \Rightarrow A^T Ax = 0 \Rightarrow x^T A^T Ax = 0 \Rightarrow (Ax)^T Ax = 0$$

$$v = Ax \Rightarrow v^T v = 0 \Rightarrow \langle v, v\rangle = 0 \Rightarrow v = 0 \Rightarrow Ax = 0 \Rightarrow x \in N_A$$

$$N_{A^T A} \subseteq N_A$$

$$N_A = N_{A^T A} \Rightarrow dim\left(N_A\right) = dim\left(N_{A^T A}\right)$$

   from the conservation of dimensions we get:

$$N = dim\left(N_A\right) + rank\left(A\right) \Rightarrow rank\left(A\right) = N - dim\left(N_A\right)$$

$$N = dim\left(N_{A^T A}\right) + rank\left(A^T A\right) \Rightarrow rank\left(A^T A\right) = N - dim\left(N_{A^T A}\right)$$

$$rank\left(A\right) = rank\left(A^T A\right)$$

   Main proof:
   $X \in R^{M\times N}$ for $M \geq N$.
   We shall prove both directions:

   (a) $\Rightarrow$ Let us assume X has full column rank: $rank\left(X\right) = N$.
   Thus, from the sentence we have just proven, $rank\left(X^T X\right) = N$.
   Note that $X^T X \in R^{N\times N}$ and is full rank, therefore $X^T X$ is invertible.
   From our prove in question 1, any solution must suffice $X^T X\theta = X^T y$.
   Since $X^T X$ is invertible, we know that there could be only is a single unique solution $\theta = \left(X^T X\right)^{-1} X^T y$.
   So, we have proven that X has full column rank $\Rightarrow$ there exists a unique solution to the linear regression.

   (b) $\Leftarrow$ Let us assume there is a uniqe solution to the linear regression.
   We shall assume towards contradiction that X is not full column rank $\left(rank\left(X\right) < N\right)$.
   Therefore, from the conservation of dimentions we know that $dim\left(N_X\right) = N - rank\left(X\right) > 0 \Rightarrow \exists v \neq 0 \ s.t. \ Xv = 0$.
   Since we assumed there exists a uniqe solution - let us denote it θ_1.

From question 1 we know that $X^T X \theta_1 = X^T y$
If we were to observe $\theta_2 = \theta_1 + v$ we can notice that

$$\begin{aligned} X^T X \theta_2 &= X^T X \left(\theta_1 + v\right) \\ &= X^T X \theta_1 + X^T X v \\ &= X^T X \theta_1 + 0 \\ &= X^T X \theta_1 = X^T y \end{aligned}$$

Again, from question 1, we know that θ_2 is also a solution because it satisfies $X^T X \theta_2 = X^T y$.
Since $v \neq 0$ we know that $\theta_2 \neq \theta_1$ and this is contradiction to the fact that there is a unique solution to the linear regression.
Thus, we know that our assumption is incorrect and X is indeed full column rank $(rank\,(X) = N)$.
So, we have proven that if there exists a unique solution to the linear regression $\Rightarrow X$ has full column rank.

So, we have proven that X has full column rank $\iff$ there exists a unique solution to the linear regression, as requested.