CS446 / ECE 449: Machine Learning, Fall 2020, Homework 1

Name: Saud Alrasheed (Sauda2)

Worked individually

Problem (1)

Solution:

1.(a): We will start by simplifying the following term,

$$L_{ols}(w) = \frac{1}{2}||Xw - y||_2^2 = \frac{1}{2}(Xw - y)^T(Xw - y) = \frac{1}{2}(w^T X^t Xw - 2w^T X^t y + y^T y)$$

We will use the fact that (Known derivatives),

$$\frac{dAw}{dw} = A, \quad and \quad \frac{dw^T Aw}{dw} = (A^T + A)w$$

for all $m \times n$ matrices. Thus,

$$\nabla L_{ols}(w) = \frac{1}{2}((X^T X + (X^T X)^T)w - 2X^T y) = X^T Xw - X^T y$$

Therefore,

$$\nabla^2 L_{ols}(w) = X^T X$$

1.(b): We will show that for any $z \in \mathbf{R^d}, z \neq 0$, we have that $z^T X^T Xz > 0$.

$$z^T X^T Xz = (Xz)^T Xz = ||Xz||_2^2$$

Thus, due to the linear independence of the columns of X, we have that,

$$||Xz||_2^2 \neq 0, \quad for\ z \neq 0$$

2.(a): We will start by simplifying the following term,

$$L_{ridge}(w) = \frac{1}{2}||Xw - y||_2^2 + \frac{\lambda}{2}||w||_2^2 = \frac{1}{2}(w^T X^t Xw - 2w^T X^t y + y^T y) + \frac{\lambda}{2}w^T w$$

Thus,

$$\nabla L_{ols}(w) = \frac{1}{2}((X^T X + (X^T X)^T)w - 2X^T y) + \lambda w = X^T Xw - X^T y + \lambda w$$

Given that $L_{ridge}(w)$ is convex, the estimator is w such that $\nabla L_{ridge}(w) = 0$,

$$\nabla L_{ols}(w) = X^T Xw - X^T y + \lambda w = 0 \implies X^T Xw + \lambda w = X^T y \implies w = (X^T Xw + \lambda I)^{-1} X^T y$$

2.(b): We will show that for any $z \in \mathbf{R^d}, z \neq 0$, we have that $z^T(X^TX + \nabla I)z > 0$.

$$z^T(X^TX + \lambda I)z = z^TX^TXz + \lambda z^Tz = ||Xz||_2^2 + \lambda ||z||_2^2$$

We have that $||Xz||_2^2 \geq 0$ and $||z||_2^2 > 0$ for any nonzero $z \in \mathbf{R^d}$ and $\lambda > 0$. Thus,

$$z^T(X^TX + \lambda I)z = ||Xz||_2^2 + \lambda ||z||_2^2 > 0$$

So, $(X^TX + \lambda I)$ is positive definite.