Introduction to Machine Learning

HOMEWORK 0

UJVAL BANGALORE UMESH
5202540

Prob 1)

1) $\underset{w}{\text{Minimize}} \|Xw - Y\|^2$ where $X \in \mathbb{R}^{m\times n}$, $y \in \mathbb{R}^{m\times 1}$, $w \in \mathbb{R}^{n\times 1}$

Given X is a full ranked matrix of size $\mathbb{R}^{m\times n}$, in order to optimize [minimize] the function $\|Xw - Y\|$ using least squares, Gradient of is used.

A vector w in $\mathbb{R}^n$, such that the Euclidean norm $\|Xw - Y\|_2^2$ is minimized is obtained by solving

$$\|Xw - Y\|_2^2 = (Xw - Y)^T (Xw - Y) =$$
$$= w^T X^T X w - w^T X^T Y - Y^T X w + Y^T Y$$
$$= w^T X^T X w - 2Y^T X w + Y^T Y$$

Taking gradient wrt w and equating it to 0.

$$\nabla_w (w^T X^T X^T w - 2Y^T X w + Y^T Y) = 0$$
$$\Rightarrow \nabla_w (w^T X^T X w) - \nabla_w (2Y^T X w) + \nabla_w Y^T Y = 0$$
$$\Rightarrow 2X^T X w - 2X^T Y = 0.$$

$2X^TXW = 2X^TY$

$\Rightarrow \quad W = (X^TX)^{-1}X^TY$

The inverse exists as X^TX is a positive definite matrix.

2) Minimize $||XW - Y||^2 + \lambda||W||^2$

Regularization is done to avoid overfitting when the hessian matrix is not positive definite

Using the result of problem 1.1, $||XW - Y||_2^2$ can be expanded as

$$W^TX^TXW - 2Y^TXW + Y^TY$$

To Minimize $||XW - Y||^2 + \lambda||W||^2$, we can take gradient and equate it to zero.

$$\nabla_W W^TX^TXW - \nabla_W 2Y^TXW + \nabla_W Y^TY + \nabla_W \lambda W^2 = 0$$

$$2X^TXW - 2X^TY + 2\lambda W = 0$$

$$W[X^TX + \lambda I] = X^TY$$

$$W = [X^TX + \lambda I]^{-1} X^TY$$

[$\because$ Inverse exists as we are adding a scalar λ to the diagonal elements of a positive definite matrix]

3) Minimize $||X\omega - Y||^2 + \lambda \omega^T L \omega$

using the expansion of $||X\omega - Y||^2$ we get.

$2x^T X\omega - 2Y^T X + Y^T Y$

To Minimize $||X\omega - Y||^2 + \lambda||\omega||^2$, we take the gradient.

$$\nabla_\omega \omega^T x^T X\omega - \nabla_\omega 2Y^T X + \nabla_\omega Y^T Y + \nabla_\omega \lambda \omega^T L \omega = 0$$

$$\Rightarrow 2x^T X\omega - 2x^T Y + 0 + \lambda(L + L^T)\omega = 0$$

$$\omega[2x^T x + \lambda(L + L^T)] = 2x^T Y$$

$$\omega = 2[2x^T x + \lambda(L + L^T)]^{-1} x^T Y$$

As L is symmetric $L = L^T$

$$\omega = [x^T x + \lambda L]^{-1} x^T Y.$$

Inverse exists as L is positive semi definite & $x^T x$ is positive definite.