Problem Set 1*

Xiaowei Zhang†

Problem 1

Yes, here we take all y, $\tilde{y}$, x in their matrix form. $\omega^* = X^+$y , Also $\tilde{\omega} = X^+\tilde{y}$, where X^+ is the Moore-Penrose pseudoinverse of X, since there is nothing change on X , X^+ will stay the same as well.

Now $\tilde{\omega} = X^+\tilde{y} = X^+$ (ay+b) = aX^+y+bX^+ = aω^*+bX^+ = g(ω^*, a, b).

So yes, we don't need to retrain on the transformed dataset y.

Problem 2

Yes, let C be a diagonal matrix of dimension of dxd with all the constant c_j from the collection of nonzero constants $c_1 ... c_d$. Suppose X is the matrix of all x from the input collection with each row be one training data x and $\overline{X}$ is the corresponding matrix with the new $\overline{x}$. ω^*in this case is $(X^TX)^{-1}X^Ty$ *and* $\overline{\omega}$ here is $(\overline{X}^T\overline{X})^{-1}\overline{X}^Ty$. By the definition of $\overline{x}$, $\overline{X}$ equals XC. Thus, $\overline{X}^T = XC^T = C^TX^T$. $\overline{\omega} = (\overline{X}^T\overline{X})^{-1}\overline{X}^Ty = (C^TX^TXC)^{-1}C^TX^Ty = C^{-1}X^{-1}(X^T)^{-1}(C^T)^{-1}C^TX^Ty = C^{-1}X^{-1}(X^T)^{-1}X^Ty = C^{-1}(X^TX)^{-1}X^Ty = C^{-1}\omega^*$. We know that C is a matrix of all the constants c_j , then C^{-1} should be as well a function of all the constants c_j. We know C is invertible because its determinant is not zero.

So yes, we can find $\overline{\omega}$ from ω^* and the collection of c_j.

Problem 3

$p(y^{(i)}|x^{(i)};\omega,\sigma_i^2) = 1/(\sigma_i\sqrt{2\pi})exp(-(y^{(i)} - f(x^{(i)};\omega))^2/(2\sigma_i^2))$.

$p(y|x;\omega,\sigma^2) = p(y^{(1)}|x^{(1)};\omega,\sigma_1^2) * p(y^{(2)}|x^{(2)};\omega,\sigma_2^2) * p(y^{(3)}|x^{(3)};\omega,\sigma_3^2) * .... * p(y^{(N)}|x^{(N)};\omega,\sigma_N^2)$.

We want to find the ω^* such that the probability would be the maximum. here we apply the same technique by introducing log function to the probability. $log^{p(y|x;\omega,\sigma^2)} = log^{1/(\sigma_1\sigma_2...\sigma_N(\sqrt{2\pi})^N)exp(-(y_1-f(x_1;\omega))^2/(2\sigma_1^2)-(y_2-f(x_2;\omega))^2/(2\sigma_2^2)...(y_N-f(x_N;\omega))^2/(2\sigma_N^2))} = -log(\sigma_1\sigma_2...\sigma_N\sqrt{2\pi}^N) - 1/(2\sigma_1^2)(y_1 - f(x_1;\omega))^2 - 1/(2\sigma_2^2)(y_2 - f(x_2;\omega))^2 - ... - 1/(2\sigma_N^2)(y_N - f(x_N;\omega))^2$ And closed form solution would be the same, i.e. $\omega^* = (X^TX)^{-1}X^Ty$

*Due: `September 23, 2021`, Student(s) worked with:

†NetID: `xz561`, Email: `xz561@rutgers.edu`