Q8

a) i) Squared error is not enough to give an unique solution especially when rank(X) < dim(w).

We use ridge regressed as it results in a robust solution.

ii) $w = (X^T X + \lambda I)^{-1} X^T X w + (X^T X + \lambda I)^{-1} X^T n$

where n is variance, I is identity and λ is a regularization parameter.

iii) We can use cross validation to select λ parameter. λ is often chosen by logarithmic spacing (1, 2, 5, 10, 20 ....)

① Split data into training & ~~test~~ validation

② Compute $w_{\lambda_i} = (A^{(t)T} A^{(t)} + \lambda_i I)^{-1} A^{(t)T} d^{(t)}$

where (t) refers to training & (v) refers to validation.

③ compute $e^2(\lambda_i) = \| A^{(V)} w_{\lambda_i} - d^{(V)} \|_2^2$

& minimize $e^2(\lambda_i)$ to get λ.

iv) A good λ would be one that controls the variance term in w.

b) If Two columns of X are identical, then $\det(X) = 0$, by definition.

Thus $X^T X$ is a singular matrix & thus is not invertible.

Thus least squares method fails when we try to calculate $(X^T X)^{-1}$.