Stat 243 PS 7

Riv Jenkins

1

We know that the standard error should be normally distributed for a linear regression. So, in the limit the standard error should approach a normally distributed variable.

2

We have the following from the notes.
$$||A||_2 = \sup_{||z||_2=1} ||Az||_2$$

We can rewrite the norm.
$$||A||_2 = \sup_{||z||_2=1} \sqrt{(Az)^T(Az)}$$

We can then decompose A with an eigendecomposition.

$$||A||_2 = \sup_{||z||_2=1} \sqrt{(\Gamma\Lambda\Gamma^T z)^T(\Gamma\Lambda\Gamma^T z)}$$

$$||A||_2 = \sup_{||z||_2=1} \sqrt{z^T\Gamma\Lambda\Gamma^T\Gamma\Lambda\Gamma^T z}$$

$$||A||_2 = \sup_{||z||_2=1} \sqrt{z^T\Gamma\Lambda^2\Gamma^T z}$$

Let $y = \Gamma^T z$. Note that $||\Gamma^T z||_2 = \sqrt{(\Gamma^T z)^T(\Gamma^T z)} = \sqrt{z^T\Gamma\Gamma^T z} = \sqrt{z^T z} = ||z||_2$. Thus if $||z||_2 = 1$, then $||y||_2 = 1$. So, we have the following.
$$||A||_2 = \sup_{||y||_2=1} \sqrt{y^T\Lambda^2 y}$$

This can be rewritten in summation notation.

$$||A||_2 = \sup_{||y||_2=1} \sqrt{\sum y_i^2\Lambda_i^2}$$

where y_i is the ith element of y and Λ_i is the ith diagonal element of Λ.

By looking at this quantity we can guess that the supremum is where $y_1 = 1$ assuming that the eigenvalues are ordered from largest magnitude to smallest. This is because y_i can be thought of as a weight for Λ_i, and if we are trying to maximize that sum, we want to give the most weight possible to the largest Λ_i. Thus we know that $||A||_2$ is the largest of the absolute values of the eigenvalues of A.

3 a)

If we perform a singular value decomposition of X we get

$$X = UDV^T$$

So we can rewrite $X^T X$.
$$X^T X = (UDV^T)^T(UDV^T)$$

$$X^T X = V D U^T U D V^T$$

Since U is an orthogonal matrix we have the following.

$$X^T X = V D^2 V^T$$

This is an eigen decomposition of $X^T X$ since V is orthogonal and D^2 satisfies the properties of an eigenvalue matrix. Also, since $D^2 \geq 0$ for each element, $X^T X$ must be positive semi-definite.

b)

For an eigenvector v and an eigenvalue λ of A, we have the following.

$$Av = \lambda v$$

We can add cv to both sides of the equation.

$$Av + cv = \lambda v + cv$$

Factor out v.

$$(A + cI)v = (\lambda + c)v$$

This implies that any eigenvector v of A is also an eigenvector of $A + cI$, and the corresponding eigenvalue for $A + cI$ is $\lambda + c$.

This means that we can compute the eigen decomposition of Σ in exactly n operations by adding Ic to the eigenvalues Λ of A. The eigenvectors are the same. So, formally:

$$\Sigma = \Gamma(\Lambda + cI)\Gamma^T$$

where $A = \Gamma \Lambda \Gamma^T$.

4 a)

```
d = t(X) * Y
partial_vector0 = C^-1 * d
partial_vector1 = (-A * partial_vector0 + b)
matrix_inverse = (A * C^-1 * t(A))^-1
beta_hat = partial_vector0 + C^-1 * (t(A) * (matrix_inverse * partial_vector1))
```

b)

```
d = t(X) %*% Y
partial_vector0 = solve(crossprod(X), d)
partial_vector1 = -A %*% partial_vector0 + b
matrix = A %*% solve(crossprod(X)) %*% t(A)
beta_hat = partial_vector0 + solve(crossprod(X), t(A) %*% (solve(matrix, partial_vector1)))
```

5 a)

The issue comes in computing $\hat{X}$. $\hat{X}$ is a matrix, so to calculate it you will need to do many matrix multiplications with large matrices. This will get very expensive computationally very quickly.

b)

By substituting in $\hat{X}$ and simplifying we have the following.

$$\hat{\beta} = (X^T Z (Z^T Z)^{-1} Z^T X)^{-1} (X^T Z (Z^T Z)^{-1} Z^T) y$$

Since we utilizing the sparsity of the matrices, this will make the problem more manageable. However, the final trick is to perform the multiplications from right to left. This will mean that the multiplactions being performed are always matrix by vector operations which are much more efficient than matrix by matrix.