# THEORETICAL QUESTIONS - ANSWERS WILLIAM LIAW

## Ordinary Least Squares

Before demonstrating that the Ordinary Least Squares (OLS) estimator has the smallest variance among all linear unbiased estimators, we first analyze the OLS estimator:

We can observe that the OLS is unbiased only if is deterministic and . In this case, assuming , we can calculate the variance of the OLS estimator:

Then, we compute the expected value and variance of the alternative estimator .

### Expected value

First, we compute the expected value of :

Thus, for to be unbiased , it is necessary that .

### Variance

Consequently, we compute the variance of :

### Conclusion

From the last expression, it’s evident that is always greater than or equal to since is non-zero, hence introduces additional variance.

The assumption of OLS that we need to use here is that is deterministic and

## Ridge regression

### Biased estimator

To show that the estimator of ridge regression, denoted as , is biased, we need to demonstrate that its expected value, , does not equal the true parameter vector .

Analyzing the ridge estimator and denoting , we have:

Now, let’s calculate the expected value of the ridge regression estimator:

Therefore, the expected value of the ridge estimator is generally different than and, consequently, biased. It can be equal to and unbiased, only if , in which case the ridge estimator becomes exactly the OLS estimator.

### SVD decomposition

Given the Singular Value Decomposition (SVD) , we can rewrite the expression for :

This solution is particularly useful when the matrix is ill-conditioned or nearly singular,. In this case, the SVD decomposition provides a numerically stable way to solve the ridge regression problem without directly inverting a potentially singular matrix. Additionally, SVD can be more computationally efficient for large datasets compared to directly computing the inverse of , which is no longer necessary through this method.

### Comparison between OLS variance and Ridge variance

First we calculate the variance of the Ridge estimator:

We know that the OLS estimator is the Ridge estimator for , thus:

Hence, it is apparent that for , the variance of the Ridge estimator is smaller than or equal to the variance of the OLS estimator, that is .

### Effect of the regularization parameter

Recalling the expression for the expected value of the Ridge estimator:

One can write the expression for the bias and variance of the ridge estimator:

Inference suggests that for small values, the ridge estimator closely resembles the OLS estimator, exhibiting low bias but high variance. Conversely, as increases, the bias of the ridge estimator amplifies in magnitude while its variance diminishes ().

### Derivation of Ridge estimator and OLS estimator expression under

As already seen on previous sections:

Assuming , these last expressions become:

Therefore:

## Elastic Net

Analyzing the Elastic Net estimator and denoting , we have:

Assuming , this last expression becomes:

Therefore: