We have observed that lasso seems to converge much faster than elastic net when $K < 10^{-3}$. For example, running the above data set with $K = 10^{-3}$ and $\tau = 0.999$ requires 140520 steps to achieve primal and dual residuals less than 10^{-7}, but lasso only takes 86 steps to achieve the same degree of convergence. We observed that the larger K is the faster is the convergence. This could be attributed to the fact that the matrix P becomes more "positive definite." Another factor is that ADMM for lasso solves an $n \times n$ linear system, but ADMM for elastic net solves a $2(n+m) \times 2(n+m)$ linear system. So even though elastic net does not suffer from some of the undesirable properties of ridge regression and lasso, it appears to have a slower convergence rate, in fact much slower when K is small (say $K < 10^{-3}$). It also appears that elastic net may be too expensive a choice if m is much larger than n. Further investigations are required to gain a better understanding of the convergence issue.

55.7 Summary

The main concepts and results of this chapter are listed below:

- Ridge regression.
- Kernel ridge regression.
- Kernel functions.
- Lasso regression.
- Elastic net regression.

55.8 Problems

Problem 55.1. Check the formula

$$(X^\top X + KI_n)^{-1} X^\top = X^\top (XX^\top + KI_m)^{-1},$$

stated in Section 55.1.

Problem 55.2. Implement the ridge regression method described in Section 55.1 in `Matlab`. Also implement ridge regression with intercept and compare solving Program (**DRR3**) and Program (**RR6′**) using centered data.

Problem 55.3. Implement the ridge regression with intercept method (**RR3b**) in `Matlab` and compare it with solving (**RR6′**) using centered data.

Problem 55.4. Verify that (**lasso3**) is equivalent to (**lasso2**) applied to the centered data $\widehat{y} = y - \overline{y}\mathbf{1}$ and $\widehat{X} = X - \overline{X}$.

Problem 55.5. Verify the fomulae obtained for the kernel ridge regression program (**KRR6′**).