یادگیری ماشین

تمرین شماره 4

Regression Algorithms

حسین توکلیان

شماره دانشجویی

## 9860571

خرداد 1399

دکتر ستار هاشمی

Questions:

⬚ Explain advantages and disadvantages of kernel ridge regression? Under what circumstances is the kernel regression better than the other methods?

　　　　Kernel methods have difficulties scaling to large modern data sets. The scalability issues are based on computational and memory requirements for working with a large matrix. In particular, the O(n^3) computational complexity for solving a single model, and the overall computational complexity associated with tuning hyperparameters are still major problems.

Ridge regression trades variance for bias. That is, the output from ridge regression is not unbiased. That's a bad thing. Ordinarily, you would not want biased estimators. However, when there is colinearity in the data, you may judge that it is worth it in order to lower the variance of those estimators.

⬚ What if λ is set to an extremely large value? Explain the role of the regularization parameter λ on training phase.

　　　　Imagine we have one line. This line suggests that for every one unit increase in weight there is one unit increase in predicted size. If the slope of the line steeper then for every one unit increase in weight the prediction for size increase by over two units. In other words when the slope of the line is steep then the prediction for size is very sensitive to relatively small changes in weights, when the slope is small then for every one unit increase in weight the prediction for size barely increases. When the slope of the line is small then the prediction for the size are much less sensitive to changes in the weights. The ridge regression penalty resulted I line that has a smaller slope when lambda is 0 the ridge regression penalty is also 0 when lambda increase the slope get smaller the prediction of the size get sensitive to weight
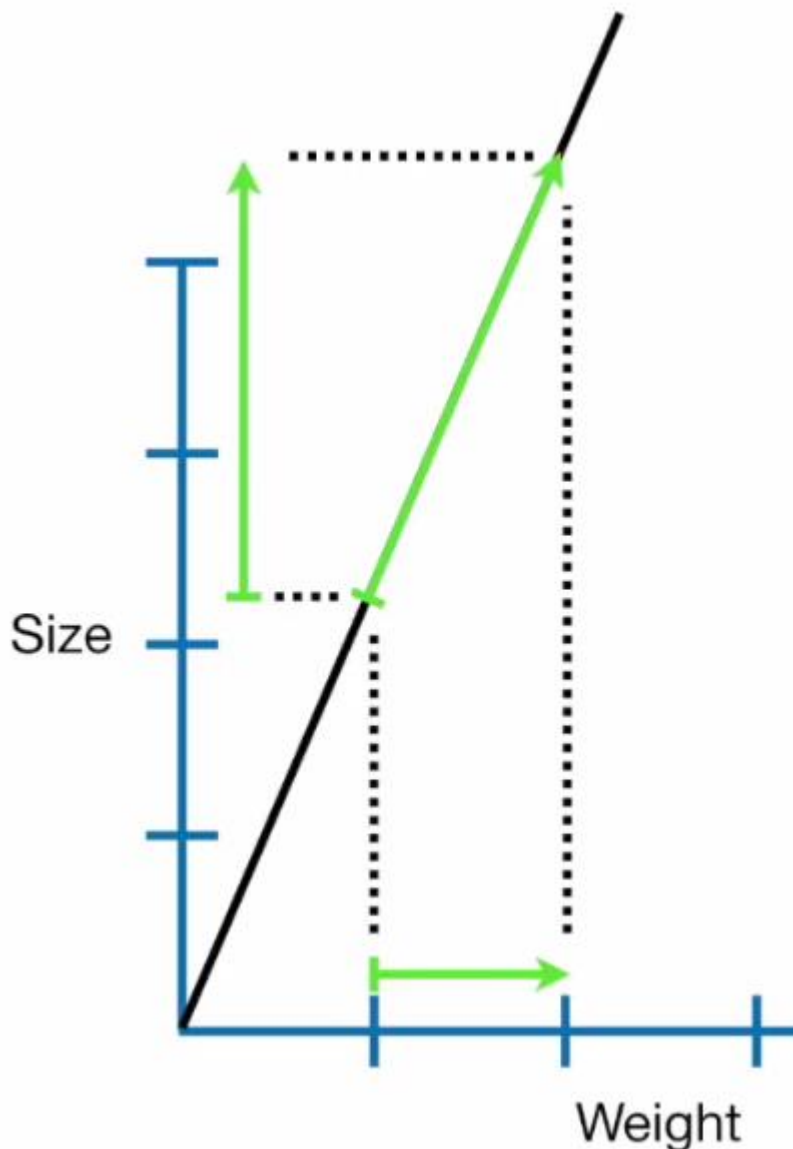
▢ In Kernel ridge regression, which kernel is better than the other, and why?

▢ When is lasso worse than ridge regression?

ridge regression can't zero out coefficients; thus, you either end up including all the coefficients in the model, or none of them. In contrast, the LASSO does both parameter shrinkage and variable selection automatically. If some of your covariates are highly correlated, you may want to look at the Elastic Net instead of the LASSO.

▢ Compare four mentioned methods, in terms of, noise, outlier, the linearly separable data, the non-linearly separable data, number of samples, and number of features. What is the effect of ε on each regression algorithm?