STAT W4400 Statistical Machine Learning Problem Set #3

Si Kai Lee sl3950@columbia.edu

March 8, 2016

Problem 1

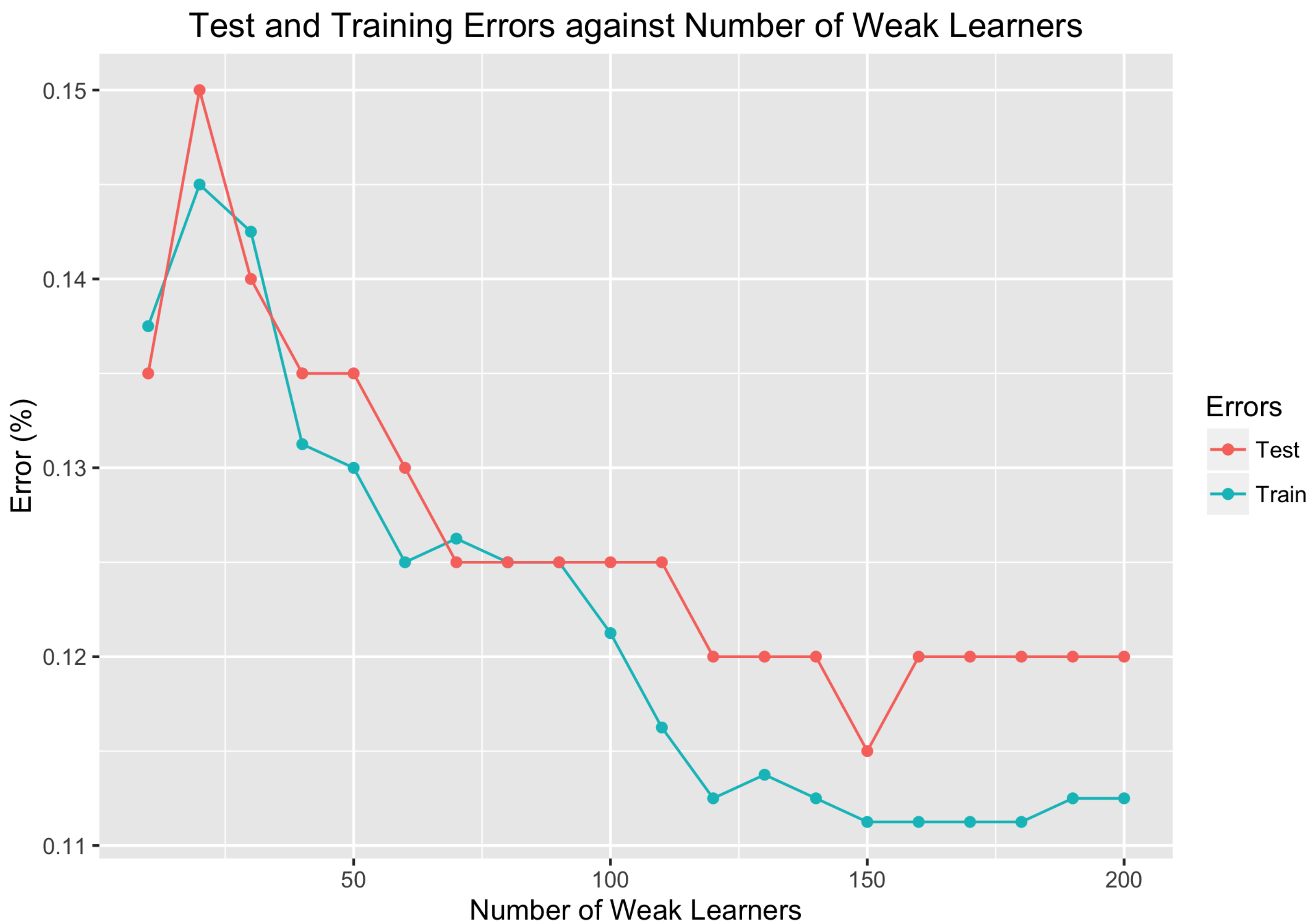

Figure 1: Training and Cross-Validated Test Errors.

Problem 2

1

The $\ell_{0.5}$ cost function encourages sparse estimates. Since the tips of the $||\beta||_{0.5}$ constraint region where $\beta_i = 0$ are closest to the contours of least squares error function $||y - X\beta||_2^2$ while the rest of the region curves away from the $||y - X\beta||_2^2$, hence the contours of $||y - X\beta||_2^2$ would most likely intersect with $||\beta||_{0.5}$ at its tips which sets the dimension β_i to 0. As the number dimensions increases, the number of tips of the constraint region increases exponentially which increases the likelihood of $||y - X\beta||_2^2$ intersecting at the tips of the constraint region i.e. dimensions being set to 0 and thus results in a sparse estimate.

On the other hand, ℓ_4 cost function does not as the $||\beta||_4$ constraint region is a rounded square which curves towards the contours of $||y - X\beta||_2^2$, therefore the two would likely intersect at a point where $\beta_i \neq 0$ so no dimensions would be turned off, leading to non-sparse estimates.

2

The geometric interpretation of minimising the $\ell_{0.5}$ cost function $\arg\min_\beta(||y - X\beta||_2^2 + \lambda||\beta||_{0.5}^{0.5})$ is equivalent to finding the contours the least squares error function $||y - X\beta||_2^2$ closest to β_{MLE} and the smallest $||\beta||_{0.5}$ constraint region. Therefore x_3 would achieve the smallest cost as it is the closest point from the contours of $||y - X\beta||_2^2$ which would intersect with $||\beta||_{0.5}$. All other points would only intersect with a larger $||\beta||_{0.5}$ constraint region.

For the second case, x_4 would achieve the smallest cost under the ℓ_4 cost function $\arg\min_\beta(||y - X\beta||_2^2 + \lambda||\beta||_4^4)$ as it is the closest point from the contours of $||y - X\beta||_2^2$ which would intersect with a smaller $||\beta||_4$ constraint region. All other points would only intersect with the larger $||\beta||_4$ shown in the figure on the right.