SDS 385 Exercise Set

Kevin Song

October 4, 2016

1 Penalized likelihood and soft thresholding

Assume our data is a gaussian distribution with $\mu = \theta$ and fixed parameter $\sigma = 1$. Then the negative log likelihood (ignoring constants) is

$$-\log\left(e^{-\frac{(y-\theta)^2}{2}}\right) = (y-\theta)^2/2 = \frac{1}{2}(y-\theta)^2$$

The next proof is a little devious—without subdifferential calculus or constrained optimization, a lot of reasoning must be applied without being able to use equations to back them up.

First, we expand the quadratic term in the expression for $S_\lambda(y)$. Since the terms which depend purely on y do not affect the minimization, we throw them out. This yields

$$S_\lambda(y) = \arg\min_\theta \frac{1}{2}y^2 - y\theta + \frac{1}{2}\theta^2 + \lambda|\theta| = \arg\min_\theta \frac{1}{2}\theta^2 - y\theta + \lambda|\theta|$$

We now take the derivative of S with respect to θ in two overlapping situations:

- If $\theta \geq 0$:
  $$S_\lambda(y) = \arg\min_\theta \frac{1}{2}\theta^2 - y\theta + \lambda\theta = \arg\min_\theta \theta^2 - 2(y-\lambda)\theta$$

  Again, we can arbitrarily add and multiply by constant terms, since the minimization is independent of constant terms. To this end, we now add $(y-\lambda)^2$ and factor to obtain the expression

  $$S_\lambda(y) = \arg\min_\theta \left(\theta - (y-\lambda)\right)^2$$

  Since we assumed $\theta > 0$, this equation has two possible solutions: if $y - \lambda$ is greater than zero, the solution is $y - \lambda$. Otherwise, the solution is zero. This is evident from the fact that the minimum of a parabola of the form $(x-a)^2$ is at a.

  In the language of cases, we find that

  $$S_\lambda(y) = \begin{cases} y - \lambda & \text{if } y - \lambda > 0 \\ 0 & \text{otherwise} \end{cases}$$

- On the other hand, if $\theta \leq 0$, we have that

  $$S_\lambda(y) = \arg\min_\theta \frac{1}{2}\theta^2 - y\theta - \lambda\theta$$

  Applying the same procedure to this equation, we find that

$$S_\lambda(y) = \arg\min_\theta \left(\theta - (y+\lambda)\right)^2$$

Which gives us

$$S_\lambda(y) = \begin{cases} y+\lambda & \text{if } y+\lambda < 0 \\ 0 & \text{otherwise} \end{cases}$$

Some further thought reveals that if $y + \lambda < 0$, y must be negative and if $y - \lambda > 0$, y must be positive. Then we have that $y + \lambda = \mathrm{sgn}(y)(|y| - \lambda)$ and $y - \lambda = \mathrm{sgn}(y)(|y| + \lambda)$. Some further thought will show that these unify in the function $S_\lambda(y) = \mathrm{sgn}(y)(|y| - \lambda)_+$

2 Nerfarious Plots

The code to generate these plots can be found in the "Ex05 Plots" iPython notebook. It contains the code for the "dummy data" tests (with known parameters), including the lambda sparsity and error plots.

3 Lasssssssso

The code to test the Python implementation of the Lasso and plot the in-sample MSE and solution paths can be found in the iPython notebook titled "LASSO."

4 Cross Validation

The code to assess cross-validation (10-fold) can be found in the notebook "Validation". This also deals with Mallow's C_p statistic. The code to do the actual cross validation is in `crossval.py`