pr623-linear-solver

April 30, 2020

1 Linear least squares solver technote

References: 1. http://www.personal.psu.edu/jhm/f90/lectures/lsq1.html 2. Boyd, S. and Vandenberghe, L., Introduction to Applied Linear Algebra, ppg 245 - 250, https://web.stanford.edu/~boyd/vmls/vmls.pdf 3. LAPACK User's Guide, Third Edition, http://www.netlib.org/lapack/lug/

We wish to fit a straight line to the data point set (x_i, y_i):

$$y_i = mx_i + b$$

where x is the the `currentCohort%year_net_uptake(z)` minus the `currentCohort%leaf_cost` and and y is the `cumulative_lai` for the cohort to the given leaf layer i (denoted as `z` in the code). Our goal is to determine the y-intercept, b, to determine what the theoretical "optimitial LAI" is that balances out the leaf layer uptake after costs (i.e. where $x = 0$).

We can model this data as a linear system:

$$\vec{y} = \mathbf{A}\vec{z}$$

where:

$$\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix}, \quad \vec{z} = \begin{bmatrix} b \\ m \end{bmatrix}$$

Here $\vec{z}$ is our solution vector consisting of the slope and intercept of our linear fit function. To determine the best fit to the set of data points, we need to minimize the residual, $\|r\|^2$:

$$\|r\|^2 = \|\mathbf{A}\vec{z} - \vec{y}\|^2$$

We arrange in this form for the `dgels` LAPACK solver:

$$\left(\mathbf{A}^T\mathbf{A}\right)\vec{z} = \mathbf{A}^T\vec{y}$$

which expands to:

$$\begin{bmatrix} N & \sum_{i=1}^{N} x_i \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i{}^2 \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} x_i y_i \end{bmatrix}$$

Given that we are retaining the code that loops through each leaf layer to maintain the original cohort method for leaf layers of less than some hardcoded minimun number (given as `nll` in the code), we can accumulate the sums and products of linear system matrices on the fly as opposed to saving off the $\vec{x}$ and $\vec{y}$ vectors using the following form:

```
nnu_clai_a(1,1) = nnu_clai_a(1,1) + 1 ! Increment for each layer used
nnu_clai_a(1,2) = nnu_clai_a(1,2) + currentCohort%year_net_uptake(z) - currentCohort%leaf_cost
nnu_clai_a(2,1) = nnu_clai_a(1,2)
nnu_clai_a(2,2) = nnu_clai_a(2,2) + (currentCohort%year_net_uptake(z) - currentCohort%leaf_cos
nnu_clai_b(1,1) = nnu_clai_b(1,1) + cumulative_lai
nnu_clai_b(2,1) = nnu_clai_b(2,1) + (cumulative_lai * &
                (currentCohort%year_net_uptake(z) - currentCohort%leaf_cost))
```