Deadline: Fr. June 21 (online) or Mo. June 23, 10:00 AM (hand-in)
Drop your printed or legible handwritten submissions into the boxes at Samelsonplatz, or upload a `.pdf` file via LearnWeb.

Remark 1. As a complementary recourse,you can read the first 2 sections (p.1-9) of the original paper. Note that there are important differences in the notation compared to the lecture slides. (see remarks)

1 Least-Angle-Regression (theory) (10 points)

A. [2p] Show that $\alpha \leq 1$ in every iteration of LARS.
Hint: Think about what happens to the correlation c_j, $j \in A_k$ as $\alpha \to 1$. How did we define the selected α?

B. [4p] Show that if $\alpha_k = 1$, then $\hat{y}_{k+1}^{\text{LARS}} = \hat{y}_{A_k}^{\text{OLS}}$, where $\hat{y}_{A_k}^{\text{OLS}}$ is the solution of the (un-regularized) ordinary least squares regression problem on the dataset (X_{A_k}, y), i.e. when using only the predictors which are in A_k.

Remark. In particular, in the last iteration of LARS, by convention we will use the stepsize $\alpha = 1$, so that the final predictor is the OLS-estimator again.

C. [4p] Show that the descent direction vector $u_k = X_{A_k} d_k$ (or equivalently $X^{(t)}\hat{\gamma}$ using the notation from the lecture slides) has the property that $\langle X_i | u_k \rangle = \langle X_j | u_k \rangle$ for all $i, j \in A_k$.
Moreover, if the data-matrix is column-normalized, that is $\|X_j\| = 1$ for all j, then u_k is even *equiangular* to all the feature vectors from the active set, i.e. $\angle(X_i, u_k) = \angle(X_j, u_k)$ for all $i, j \in A_k$

Remark 2. This is why the method is called least angle regression – if the vector wasn't equiangular, then at least one of the angles would be bigger.)

2 Least-Angle-Regression (practice) (10 points)

A. [10p] Perform 3 iterations of LARS on dataset from table 1. Start with $\beta = 0$ and update using the steplength formula. In the last iteration, use $\alpha = 1$.

Table 1

x_1	x_2	x_3	y
−1	−1	0	−1
0	1	2	0
1	−1	−2	3
2	2	1	3

Remark 3. In the slides, γ is used as the direction, we use u instead. In the paper, γ denotes the step size. We call it α instead. Use the following notation:

$$
\begin{aligned}
A_k &= A_{k-1} \cup \operatorname*{argmax}_{j \in A_{k-1}^c} |c_{k-1}| && \text{(active set)} \\
A_k^c &= \{1, \ldots, m\} \setminus A_k && \text{(complement of active set)} \\
\hat{y}_k &= X_{A_k} \beta_{A_k}^{(k)} && \text{(only take entries with index in active set)} \\
\beta_{A_k}^{(k+1)} &= \beta_{A_k}^{(k)} + \alpha_k d_k && \text{(update selected entries)} \\
\hat{y}_{k+1} &= \hat{y}_k + \alpha_k u_k \\
r_k &= y - \hat{y}_k && \text{(residual)} \\
c &= X_{A_k}^T r_k && \text{(correlation terms)} \\
\hat{c} &= \max |c| \\
d_k &= (X_{A_k}^\intercal X_{A_k})^{-1} X_{A_k}^T r_k \\
u_k &= X_{A_k} d_k \\
\alpha_k &= \min_{j \in A_k^c}{}^{+} \left\{ \frac{\hat{c} - c_j}{\hat{c} - X_j^\intercal u_k}, \frac{\hat{c} + c_j}{\hat{c} + X_j^\intercal u_k} \right\}
\end{aligned}
$$

Note that in the paper, u is re-scaled by a factor of $\frac{A_{\mathcal{A}}}{\hat{C}}$, hence the last formula looks different. Crucially, αu is the same, whether or not one does the re-scaling.