

## Assignment 7

---

### 1 Adding a new feature in least-squares regression

Let  $RSS(w) = \|Xw - y\|_2^2$  be the residual sum of squares.

a. Show that

$$\begin{aligned}\frac{\partial}{\partial w_k} RSS(w) &= a_k w_k - c_k \\ a_k &= 2 \sum_{i=1}^n x_{ik}^2 = 2 \|x_{:,k}\|^2 \\ c_k &= 2 \sum_{i=1}^n x_{ik} (y_i - w_{-k}^T x_{i,-k}) = 2 x_{:,k}^T r_k\end{aligned}$$

where  $w_{-k} = w$  without component  $k$ ,  $x_{i,-k}$  is  $x_i$  without component  $k$ , and  $r_k = y - w_{-k}^T x_{:, -k}$  is the residual due to using all the features except feature  $k$ . Hint: Partition the weights into those involving  $k$  and those not involving  $k$ .

b. Show that if  $\frac{\partial}{\partial w_k} RSS(w) = 0$ , then

$$\hat{w}_k = \frac{x_{:,k}^T r_k}{\|x_{:,k}\|^2}$$

Hence when we sequentially add features, the optimal weight for feature  $k$  is computed by computing orthogonally projecting  $x_{:,k}$  onto the current residual.