Statistics 312/612, fall 2016
Homework # 9
Due: Wednesday 16 November

[1] Suppose A is a $k \times n$ matrix of rank m, with svd

$$A = UDV^T = \sum\nolimits_{i \le m} d_i u_i v_i^T .$$

Remember that $\{u_i : i \in [k]\}$ is an onb for $\mathbb{R}^k$ and $\{v_j : j \in [n]\}$ is an onb for $\mathbb{R}^n$. Suppose we are given some vector z in $\mathbb{R}^k$ that is known to belong to the column space of A, that is, $z = AF$ for some given F in $\mathbb{R}^n$. Suppose we want to represent z as AA^TG for some G in $\mathbb{R}^k$ (cf. Tibshirani, 2013, page 1461).

(i) (10 points) Show that the general solution to $AF = AA^TG$ is

$$G = UD^{-1}V^TF + g \qquad \text{with } g \in \{w \in \mathbb{R}^k : A^Tg = 0\}.$$

(ii) (10 points) If $m = k$, explain why AA^T is non-singular, so that $G = (AA^T)^{-1}AF$ is the unique solution.

[2] Suppose f is a real-valued function on $\mathbb{R}^k$ defined by $f(z) = \sum_{i \in [k]} \psi_i(z_i)$, where $z = (z_1, \dots, z_k)$ and each ψ_j is a convex function on the real line.

(i) (5 points) Show that f is convex.

(ii) (5 points) If each ψ_j is strictly convex, show that f is also strictly convex.

(iii) (extra credit) For each $\alpha > 1$ show that the function $\psi(t) = |t|^\alpha$ is strictly convex on the real line. Hint: Show that ψ is differentiable with strictly increasing derivative.

[3] Suppose $y \in \mathbb{R}^n$ and $X = (x_1, \dots, x_p)$ is an $n \times p$ matrix whose columns are unit vectors: $\|x_j\|_2 = 1$ for each j. For a fixed $\lambda > 0$ define

$$G(b) = G_\lambda(b) = Q(b) + \lambda \|b\|_1 \qquad \text{where } Q(b) = \tfrac{1}{2} \|y - Xb\|_2^2 .$$

The directional derivative at b of G in the direction u is defined as

$$D_G(b, u) = \lim_{t \downarrow 0} \left(f(b + tu) - f(b) \right) / t.$$

Convexity ensures that a vector $\widehat{b}$ minimizes G if and only if $D_G(\widehat{b}, u) \ge 0$ for every direction u.

(i) (10 points) Show that

$$Q(\widehat{b} + tu) - Q(\widehat{b}) = -t\langle y - Xb, Xu \rangle + o(|t|) \qquad \text{as } t \to 0.$$

Deduce that Q has directional derivative

$$D_Q(\widehat{b}, u) = -\sum\nolimits_{j \in [p]} u_j x_j^T (y - X\widehat{b}).$$

(ii) (10 points) Remember that the convex function $\psi(t) = |t|$ on the real line has right- and left-derivatives

$$\mathcal{R}(t) = \mathbb{1}\{t \ge 0\} - \mathbb{1}\{t < 0\} = \text{sgn}(t) + \mathbb{1}\{t = 0\}$$
$$\mathcal{L}(t) = \mathbb{1}\{t > 0\} - \mathbb{1}\{t \le 0\} = \text{sgn}(t) - \mathbb{1}\{t = 0\}$$

where

$$\text{sgn}(t) = \mathbb{1}\{t > 0\} - \mathbb{1}\{t < 0\} = \begin{cases} +1 & \text{if } t > 0 \\ 0 & \text{if } t = 0 \\ -1 & \text{if } t < 0 \end{cases} .$$

Explain why

$$\psi(\widehat{b}_j + tu_j) - \psi(\widehat{b}_j) = tu_j \left(\mathcal{R}(\widehat{b}_j)\mathbb{1}\{u_j > 0\} + \mathcal{L}(\widehat{b}_j)\mathbb{1}\{u_j < 0\} \right) + o(|t|) \qquad \text{as } t \to 0.$$

Deduce that the function $h(b) = \|b\|_1$ has directional derivatives

$$D_h(\widehat{b}, u) = \sum\nolimits_{j \in [p]} u_j \left(\mathcal{R}(\widehat{b}_j)\mathbb{1}\{u_j > 0\} + \mathcal{L}(\widehat{b}_j)\mathbb{1}\{u_j < 0\} \right).$$

(iii) (5 points) Explain why $\widehat{b}$ minimizes G if and only if

$$\sum\nolimits_{j\in[p]} u_j \left(\lambda\mathcal{R}(\widehat{b}_j)\mathbb{1}\{u_j > 0\} + \lambda\mathcal{L}(\widehat{b}_j)\mathbb{1}\{u_j < 0\} - x_j^T(y - X\widehat{b})\right) \ge 0 \qquad \text{for every } u.$$

(iv) (5 points) Explain why $\widehat{b}$ minimizes G if and only if the following inequalities hold for every j:

$$\lambda\mathcal{R}(\widehat{b}_j) - x_j^T(y - X\widehat{b}) \ge 0$$
$$-\lambda\mathcal{L}(\widehat{b}_j) + x_j^T(y - X\widehat{b}) \ge 0$$

(v) (10 points) Explain why $\widehat{b}$ minimizes G if and only if

$$\begin{aligned} x_j^T(y - X\widehat{b}) &= \lambda && \text{for all } j \text{ where } \widehat{b}_j > 0 \\ x_j^T(y - X\widehat{b}) &= -\lambda && \text{for all } j \text{ where } \widehat{b}_j < 0 \\ |x_j^T(y - X\widehat{b})| &\le \lambda && \text{for all } j \text{ where } \widehat{b}_j = 0 \end{aligned}$$

[4] Use the same notation as in Problem [3]. For each $\lambda \ge 0$ suppose $\widehat{b}(\lambda)$ minimizes G_λ. Define $m(\lambda) = G_\lambda(\widehat{b}(\lambda))$ and $q(\lambda) = Q(\widehat{b}(\lambda))$ and $\ell(\lambda) = \left\|\widehat{b}(\lambda)\right\|_1$. Consider any pair of λ values: $0 \le \lambda_1 < \lambda_2$. Abbreviate $\widehat{b}(\lambda_i)$ to a_i and define $\delta = \lambda_2 - \lambda_1$.

(i) (10 points) Explain why

$$m(\lambda_1) = G_{\lambda_1}(a_1) \le G_{\lambda_1}(a_2) \le G_{\lambda_2}(a_2) = m(\lambda_2).$$

That is, explain why $m(\lambda)$ increases as λ increases.

(ii) (10 points) Explain why

$$Q(a_1) + \lambda_1\|a_1\|_1 \le Q(a_2) + \lambda_1\|a_2\|_1$$
$$Q(a_2) + \lambda_2\|a_2\|_1 \le Q(a_1) + \lambda_2\|a_1\|_1 .$$

Deduce that $\|a_2\|_1 \le \|a_1\|_1$. That is, $\ell(\lambda)$ decreases as λ increases. Hint: Add.

(iii) (10 points) Explain why $Q(\lambda)$ increases as λ increases.

(iv) (extra credit) Use the diabetes data from the LARS paper (in **R**: `data(diabetes)`) and the output from `out <- lars(db$x,db$y,type="lasso")` to draw plots of $m(\lambda)$, $q(\lambda)$, and $\ell(\lambda)$ versus λ. Show your code.

References

Tibshirani, R. J. (2013). The lasso problem and uniqueness. *Electron. J. Statist. 7*, 1456–1490.