P9120 - hw1

1. To prove the Bridge (β) function is convex, we need to use the first and second order charaterization of convex function:

First we seperate the Bridge function into two parts:

Let $F(\beta) = (y - X\beta)^T (y - X\beta)$ be the first part,

$$G(\beta) = \lambda \sum_{j=1}^{p} |\beta_j|^q = \sum_{j=1}^{p} |\beta_j|_q^q$$

Expand $F(\beta) = y^T y - y^T X\beta - \beta^T X^T y + \beta^T X^T X\beta$

$$\Rightarrow \frac{\partial F(\beta)}{\partial \beta} = -X^T y - X^T y + 2X^T X\beta = -2X^T y + 2X^T X\beta$$

$$\Rightarrow \frac{\partial F(\beta)}{\partial \beta^2} = 2X^T X \geq 0$$ $\Rightarrow$ It's a positive semi-definite matrix. $\Rightarrow F(\beta)$ is convex.

By definition: $F(\beta)$ is convex since second-order $\nabla^2 F(\beta) \geq 0$ for all $\beta \in dom(f)$ $\Leftrightarrow$ $F(\beta)$ is convex.

Then we turn to $G(\beta)$, for any vector β_1, β_2.

$\| t\beta_1 + (1-t)\beta_2 \|_q \leq t\|\beta_1|_q + (1-t)\|\beta_2\|_q \quad (\forall t \in [0,1])$ [Minkovski Inequality]

By definition $|\beta|_q$ is convex if $q \geq 1$

So $G(\beta) = \lambda |\beta|_q^q$ is then rewritten as.

$\lambda \| t\beta_1 + (1-t)\beta_2 \|_q^q \leq \lambda t \|\beta_1\|_q^q + \lambda(1-t)\|\beta_2\|_q^q$, thus, we derive that

$G(\beta)$ is also convex if $\lambda > 0$ and $q \geq 1$, since $(m^q \geq n^q$ for $m, n > 0$ and $q \geq 1)$

Thus, $Bridge_\lambda(\beta) = H(\beta) + G(\beta)$ is a convex function when $\lambda > 0$ and $q \geq 1$

Besides, for $q \neq 1$, the equal sign cannot be attained as below shown:

$$\lambda \| t\beta_1 + (1-t)\beta_2 \|_q^q < \lambda t \|\beta_1\|_q^q + \lambda(1-t)\|\beta_2\|_q^q$$

So, we conclude that $Bridge_\lambda(\beta) = H(\beta) + G(\beta)$ is strictly Convex when $\lambda > 0$, $q > 1$

To sum up: Bridge (β) is convex when $\lambda > 0$, $p \geq 1$.

Bridge (β) is strictly convex when $\lambda > 0$, $p > 1$.

(b) From Question (a), we know that:

① The bridge function is a convex function in β, and the second-order derivative proofs to be positive, So the minimum over the domain (f) can be attained, thus, the bridge function has at least one solution, there is minimizer for the function.

② However, to prove its uniqueness, we need to prove: if the is another β_m, $Bridge(\beta_m) = Bridge(\beta_n)$ in which m, n allows $\arg\min Bridge(\beta)$, By definition of strictly convex: For $t \in (0,1)$,

$$Bridge(t\beta_m + (1-t)\beta_n) < t Bridge(\beta_m) + (1-t) Bridge(\beta_n)$$
$$= t Bridge(\beta_m) + Bridge(\beta_n) - t Bridge(\beta_n)$$
$$= Bridge(\beta_n)$$

Since β_n is set to attain the min value, the above function can not hold, so the minimizer should be unique.

For $\sum_{j=1}^{p} |\hat{\beta}_j(\lambda)|^q \le t_0$, we are assuming $\tilde{\beta}(\lambda)$ is the minimizer for $F(X) = (y - X\beta)^T(y - X\beta)$

So, for Bridge minimizer $\hat{\beta}(\lambda) \in dom(f)$:

$$(y - X\tilde{\beta}(\lambda))^T(y - X\tilde{\beta}(\lambda)) + \|\tilde{\beta}(\lambda)\|_q^q \le (y - X\hat{\beta}(\lambda))^T(y - X\hat{\beta}(\lambda)) + \|\hat{\beta}(\lambda)\|_q^q$$

(Assume $\|\tilde{\beta}(\lambda)\|_q^q \le \|\hat{\beta}(\lambda)\|_q^q$)

Then the equation suggests that $\hat{\beta}(\lambda)$ could not be the minimizer of $Bridge(\beta)$ as there is another $\tilde{\beta}(\lambda)$ appears to have lower value.

So, the part of $\sum_{j=1}^{p} |\hat{\beta}_j(\lambda)|^q \le t_0$.

(c) The Bridge function is proved to be convex;

Suppose there're 2 minimizers $\hat{\beta}_1(\lambda)$ and $\hat{\beta}_2(\lambda)$.

for $\forall t \in (0,1)$ M denotes the min value obtained by $\hat{\beta}_1(\lambda)$ and $\hat{\beta}_2(\lambda)$ under the LASSO regression.

$$\|y - X[t\hat{\beta}_1(\lambda) + (1-t)\hat{\beta}_2(\lambda)]\|_2^2 + \lambda\|t\hat{\beta}_1(\lambda) + (1-t)\hat{\beta}_2(\lambda)\| < tM + (1-t)\cdot M = M.$$

Where the equal sign can not be achieved as the function is strict convex.

this is contradict to the Assumption M is the minimum value as a lower value was attained.

Thus, two minimizers should be equal and the panelty function should take the same value.

$$S(\lambda) \triangleq \sum_{j=1}^{p} |\hat{\beta}_j(\lambda)|^q$$

While for proving $S(\lambda) \le t_0$, let $\hat{\beta}(\lambda)$ be the minimizer of Bridge function and $\tilde{\beta}(\lambda)$ be the minimizer of $F(\beta) = (y - X\beta)^T(y - X\beta)$

Suppose $\|\hat{\beta}(\lambda)\| > \|\tilde{\beta}(\lambda)\|$ holds,

$$Bridge_\lambda(\tilde{\beta}) = (y - X\tilde{\beta}(\lambda))^T(y - X\tilde{\beta}(\lambda)) + \|\tilde{\beta}(\lambda)\|$$
$$< (y - X\hat{\beta}(\lambda))^T(y - X\hat{\beta}(\lambda)) + \|\hat{\beta}(\lambda)\|$$

Cause $\hat{\beta}(\lambda)$ cannot be the unique minimizer, so there's some contradiction.

$$\Rightarrow S(\lambda) \triangleq \sum_{j=1}^{p} |\hat{\beta}_j(\lambda)|^q \le t_0$$

(d) Since $G(\beta)$ is a convex function,

so, $G(\beta)$ has at least one subgradient at

every point in relint dom G, also G is differentiable $\nabla G(a)$ is subgradient of G at a.

we can define vector h as a subgradient at any β, and it satisfies that:

$$g(\beta') \geq g(\beta) + h^T(g(\beta') - g(\beta)) \quad \text{for } \forall \beta \in \text{dom}(\beta)$$

hence: $g(\beta') \leq g(\beta) \Rightarrow h^T(g(\beta') - g(\beta)) \leq 0$, $\partial g(\beta)$ denotes the subdifferential that includes all possible subgradients of g at β.

① To prove Condition in (d) $\Rightarrow$ Condidition in (c)

The lagrangian form of P_2:

$$L(\beta, r) = (y - X\beta)^T(y - X\beta) + r(\|\beta\|_q^q - s(\lambda))$$

$\Rightarrow$ the lagrangian condition: $0 \in -2X^Ty + 2X^TX\beta + r\partial G(\beta)$

② To prove Condition in (c) $\Rightarrow$ Condition in (d).

Due to the constraints specified, we checked:

for $(y - X\beta)^T(y - X\beta)$ subject to $\sum_{j=1}^{p} |\beta_j|^q \leq s(\lambda)$ ... (1)

a. $\|\beta\|_q^q - s(\lambda) \leq 0$ holds as β can be all zeros.

b. When $\lambda = r$, $\beta = \beta(r)$, the complementary condition is satisfied at solution point β.

To conclude, to minimize Bridge(β) in (c) is equvalent to that minimizing $(y - X\beta)^T(y - X\beta)$ with $\sum_{j=1}^{p} |\beta_j|^q \leq s(\lambda)$.