HW5_561

October 14, 2018

0.1 Jingrong Tian

1 Since $(f \circ \gamma)'(t) = \nabla f(\gamma(t))^T \gamma'(t) = \frac{\partial f}{\partial x}(\gamma(t))\gamma_1'(t) + \frac{\partial f}{\partial y}(\gamma(t))\gamma_2'(t)$
we can get $\nabla(f \circ g)(x,y) = \frac{\partial f}{\partial x}(g(x,y))g_1'(x,y) + \frac{\partial f}{\partial y}(g(x,y))g_2'(x,y)$
$= \begin{pmatrix} g_1'(x,y) & g_2'(x,y) \end{pmatrix} \begin{pmatrix} \frac{\partial f}{\partial x}(g(x,y)) \\ \frac{\partial f}{\partial y}(g(x,y)) \end{pmatrix}$
$= \begin{pmatrix} g_1'(x,y) \\ g_2'(x,y) \end{pmatrix}^T \begin{pmatrix} \frac{\partial f}{\partial x}(g(x,y)) \\ \frac{\partial f}{\partial y}(g(x,y)) \end{pmatrix}$
$= Dg(x,y)^T \nabla f(g(x,y))$

2.(a) Part 1 We denote, $y = \begin{pmatrix} y_1 \\ y_2 \\ ... \\ y_{10} \end{pmatrix}$,
$x = \begin{pmatrix} 1, x_1 \\ 1, x_2 \\ ... \\ 1, x_{10} \end{pmatrix}$,
$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$
The original function then becomes

$$f(\beta) = \frac{1}{10}(y - x\beta)^{-1}(y - x\beta)$$

$$= \frac{1}{10}(y^{\mathrm{T}}y - 2y^{\mathrm{T}}x\beta + \beta^{\mathrm{T}}x^{\mathrm{T}}x\beta)$$

$$\frac{\partial f}{\partial \beta_0}(\beta) = \frac{1}{10}(-2(1,0)x^{\mathrm{T}}y + 2(1,0)x^{\mathrm{T}}x\beta)$$

$$= \frac{2}{10}((1,0)(x^{\mathrm{T}}x\beta - x^{\mathrm{T}}y)) = (1,0)\frac{1}{5}(x^{\mathrm{T}}x\beta - x^{\mathrm{T}}y)$$

$$\frac{\partial f}{\partial \beta_1} = (0,1)\frac{1}{5}(x^{\mathrm{T}}x\beta - x^{\mathrm{T}}y)$$

$$\nabla f = \frac{1}{5}(x^{\mathrm{T}}x\beta - x^{\mathrm{T}}y)$$

$$(\frac{\partial}{\partial \beta_0})\nabla f = \frac{1}{5}(x^{\mathrm{T}}x \begin{pmatrix} 1 \\ 0 \end{pmatrix})$$

$$(\frac{\partial}{\partial \beta_0})\nabla f = \frac{1}{5}(x^{\mathrm{T}}x \begin{pmatrix} 0 \\ 1 \end{pmatrix})$$

$$\nabla^2 f = \frac{1}{5}\begin{pmatrix} 10, 1 \\ 1, 7 \end{pmatrix}$$

Since the determinant is greater than zero and hessian has all its eigenvalues as positive, thus hessian is positive definite. Therefore, the function is strictly convex.

2a Part 2 The function is strictly convex. As proved by the sufficiant condition for comvexity, this objective has a unique minizer $({\beta_0}^*, {\beta_1}^*)$ when $\frac{\partial f}{{\beta_0}^*} = \frac{\partial f}{{\beta_1}^*} = 0$

$$\nabla f = \frac{1}{5}(x^{\mathrm{T}}x\beta - x^{\mathrm{T}}y) = \mathbf{0}$$

Gives the result $({\beta_0}^*, {\beta_1}^*) = (\frac{-5}{69}, \frac{50}{69})$

2(b) Part 1 Define $x = (x_1, x_2, ...x_{10})$ $y = (y_1, y_2, ...y_{10})$ $\beta = (\beta_0, \beta_1)$

$$f(\beta) = -logP(y|X, \beta) = \sum_{i=1}^{10} log(1 + exp(-y_i x_i \beta))$$

To show that the negative log like function is convex, we consider the partial derivatives. Define $g(z) = \frac{1}{1+e^{-z}}$ (logit function). Note that $1 - g(z) = \frac{e^{-z}}{1+e^{-z}}$ and $\frac{\partial g(z)}{\partial z} = -g(z)(1 - g(z))$

$$\frac{\partial f(\beta)}{\partial \beta_i} = -\sum_{i=1}^{10} y_i x_i (1 - g(y_i \beta^{\mathrm{T}} x_i))$$

$$\frac{\partial^2 f(\beta)}{\partial \beta_i \beta_j} = \sum_{i=1}^{10} {y_i}^2 x_i g(y_i \beta^{\mathrm{T}} x_i)(1 - g(y_i \beta^{\mathrm{T}} x_i))$$

To show that the objective is convex, we first show that the Hession is PSD. Choose any vector a(dimension as 1*2).

Denote $p_i = g(y_i \beta^{\mathrm{T}} x_i)(1 - g(y_i \beta^{\mathrm{T}} x_i))$

$$a^{\mathrm{T}} \nabla^2 a = \sum_{i=1}^{10} a^{\mathrm{T}} p_i {p_i}^{\mathrm{T}} a^{\mathrm{T}} \geq 0$$

Since $a^{\mathrm{T}} p_i {p_i}^{\mathrm{T}} a^{\mathrm{T}} = (a^{\mathrm{T}} p_i)^2$

Hence the Hessian is PSD.

While β is L2 regularized in this question $c = \frac{1}{10}$ is added as a regulazer, the hessian therefore is positive definite and hence the objective is strictly convex.

2b Part(2) Sufficiant Condition If the following two equations hold,

$$\frac{\partial f}{\partial \beta_0} = \frac{1}{10}(1 + exp(-y_i\beta_1 x_i)exp(-y_i\beta_0))^{-1}(-y_i)exp(-y_i\beta_0)\begin{pmatrix}1\\0\end{pmatrix} = 0$$

$$\frac{\partial f}{\partial \beta_0} = \frac{1}{10}(1 + exp(-y_i\beta_1 x_i)exp(-y_i\beta_0))^{-1}(-y_i x_i)exp(-y_i\beta_1 x_i)\begin{pmatrix}0\\1\end{pmatrix} = 0$$

Then the solution vector to these two equations β=(${\beta_0}^*$,${\beta_1}^*$) is the minimizer to the objective

Necessary Condition If $({\beta_0}^*, {\beta_1}^*)$ is known to be the minimizer to this obejctive, then the following two equations naturally holds

$$\frac{\partial f}{\partial \beta_0} = \frac{1}{10}(1 + exp(-y_i{\beta_1}^* x_i)exp(-y_i{\beta_0}^*))^{-1}(-y_i)exp(-y_i{\beta_1}^*)\begin{pmatrix}1\\0\end{pmatrix} = 0$$

$$\frac{\partial f}{\partial {\beta_0}^*} = \frac{1}{10}(1 + exp(-y_i{\beta_1}^* x_i)exp(-y_i{\beta_0}^*))^{-1}(-y_i x_i)exp(-y_i{\beta_1}^* x_i)\begin{pmatrix}0\\1\end{pmatrix} = 0$$

3

$$A = \begin{pmatrix}a, b\\b, c\end{pmatrix}$$

$$x = (x, y)^T$$

$$f(x, y) = max\frac{1}{2}x^{\mathrm{T}}Ax = \frac{1}{2}(ax^2 + 2bxy + cy^2)$$

constrain: $g(x, y) = x^2 + y^2 - 1$

$$\nabla f(x, y) = \begin{pmatrix}ax + by\\bx + cy\end{pmatrix} = \begin{pmatrix}a, b\\b, c\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix} = Ax$$

$$\nabla g(x, y) = 2\begin{pmatrix}x\\y\end{pmatrix}$$

$$\begin{pmatrix}a, b\\b, c\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix} = 2\lambda\begin{pmatrix}x\\y\end{pmatrix}$$

Thus x is both the eigenvector of A and the solution to f(x,y) subject to $x^2 + y^2 = 1$

```
In [ ]:
```