ANLY561 Assignment 5

Hongyang Zheng

1.

$$
\begin{aligned}
\nabla(f \circ g)(x, y) &= \begin{pmatrix} \frac{\partial(f \circ g)}{\partial x}(x, y) \\ \frac{\partial(f \circ g)}{\partial y}(x, y) \end{pmatrix} \\
&= \begin{pmatrix} \nabla f(g(x, y))^T \partial_1 g(x, y) \\ \nabla f(g(x, y))^T \partial_2 g(x, y) \end{pmatrix} \\
&= \begin{pmatrix} \begin{pmatrix} \frac{\partial(f(g))}{\partial x}(x, y) \\ \frac{\partial(f(g))}{\partial y}(x, y) \end{pmatrix}^T \begin{pmatrix} \frac{\partial g_1}{\partial x}(x, y) \\ \frac{\partial g_2}{\partial x}(x, y) \end{pmatrix} \\ \begin{pmatrix} \frac{\partial(f(g))}{\partial x}(x, y) \\ \frac{\partial(f(g))}{\partial y}(x, y) \end{pmatrix}^T \begin{pmatrix} \frac{\partial g_1}{\partial y}(x, y) \\ \frac{\partial g_2}{\partial y}(x, y) \end{pmatrix} \end{pmatrix} \\
&= \begin{pmatrix} \frac{\partial(f(g))}{\partial x}(x, y)\frac{\partial g_1}{\partial x}(x, y) + \frac{\partial(f(g))}{\partial y}(x, y)\frac{\partial g_2}{\partial x}(x, y) \\ \frac{\partial(f(g))}{\partial x}(x, y)\frac{\partial g_1}{\partial y}(x, y) + \frac{\partial(f(g))}{\partial y}(x, y)\frac{\partial g_2}{\partial y}(x, y) \end{pmatrix} \\
&= \begin{pmatrix} \frac{\partial g_1}{\partial x}(x, y) & \frac{\partial g_2}{\partial x}(x, y) \\ \frac{\partial g_1}{\partial y}(x, y) & \frac{\partial g_2}{\partial y}(x, y) \end{pmatrix} \begin{pmatrix} \frac{\partial f(g)}{\partial x}(x, y) \\ \frac{\partial f(g)}{\partial y}(x, y) \end{pmatrix} \\
&= Dg(x, y)^T \nabla f(g(x, y))
\end{aligned}
$$

2.

(a)

$$
\begin{aligned}
\ell_{lin}(\beta_0, \beta_1) =& \frac{1}{10}(3(-1 + \beta_1 - \beta_0)^2 + 2(1 - \beta_0)^2 + (-1 - \beta_0)^2 + 3(1 - \beta_1 - \beta_0)^2 + (-1 - \beta_1 - \beta_0)^2) \\
=& \frac{1}{10}(7\beta_1^2 + 10\beta_0^2 + 2\beta_1\beta_0 - 10\beta_1 + 10)
\end{aligned}
$$

Let

$$
N = \nabla^2 \ell_{lin}(\beta_0, \beta_1) = \frac{1}{10}\begin{pmatrix} 20 & 2 \\ 2 & 14 \end{pmatrix}
$$

$det(N) = \frac{1}{10}(280 - 4) = 27.6 > 0$ and N is a symmetric matrix.

Therefore, N is positive definite and $\ell_{lin}(\beta_0, \beta_1)$ is strictly convex.

$\partial_1 = \frac{1}{10}(20\beta_0 + 2\beta_1)$, $\partial_2 = \frac{1}{10}(14\beta_1 + 2\beta_0 - 10)$

Let $\partial_1 = 0$ and $\partial_2 = 0$, we have $\beta_0 = -0.07246377, \beta_1 = 0.72463768$

Since $\ell_{lin}(\beta_0, \beta_1)$ is strictly convex, it only has one minimizer, which is $(\beta_0, \beta_1) = (-0.07246377, 0.72463768)$

(b)

Let

$$f(\beta_0,\beta_1) = log(1+e^{-y_i(\beta_1 x_i+\beta_0)})$$

and let

$$A = e^{-y_i(\beta_1 x_i+\beta_0)}$$

$$\partial_1 f = \frac{A * -y_i}{1+A}, \partial_2 f = \frac{A * -y_i x_i}{1+A}$$

$$\partial_{1,1} f = \frac{(A * -y_i)(-y_i)(1+A) - (A * -y_i)(A * -y_i)}{(1+A)^2} = \frac{Ay_i^2}{(1+A)^2}$$

$$\partial_{1,2} f = \frac{(A * -y_i x_i)(-y_i)(1+A) - (A * -y_i)(A * -y_i x_i)}{(1+A)^2} = \frac{Ay_i^2 x_i}{(1+A)^2}$$

$$\partial_{2,1} f = \frac{(A * -y_i)(-y_i x_i)(1+A) - (A * -y_i x_i)(A * -y_i)}{(1+A)^2} = \frac{Ay_i^2 x_i}{(1+A)^2}$$

$$\partial_{2,2} f = \frac{(A * -y_i x_i)(-y_i x_i)(1+A) - (A * -y_i x_i)(A * -y_i x_i)}{(1+A)^2} = \frac{Ay_i^2 x_i^2}{(1+A)^2}$$

Let

$$A = \frac{3e^{-\beta_1+\beta_0}}{1+e^{-\beta_1+\beta_0}}, B = \frac{2e^{-\beta_0}}{1+e^{-\beta_0}}, C = \frac{e^{\beta_0}}{1+e^{\beta_0}}, D = \frac{3e^{-\beta_1-\beta_0}}{1+e^{-\beta_1-\beta_0}}, E = \frac{e^{\beta_1+\beta_0}}{1+e^{\beta_1+\beta_0}}$$

All the terms are greater than 0

Then plug $x_i's, y_i's$, we have

$$\partial_{1,1}\ell_{log}(\beta_0,\beta_1) = \frac{1}{10}(A+B+C+D+E) > 0$$

$$\partial_{1,2}\ell_{log}(\beta_0,\beta_1) = \partial_{2,1}\ell_{log}(\beta_0,\beta_1) = \frac{1}{10}(-A+D+E) > 0$$

$$\partial_{2,2}\ell_{log}(\beta_0,\beta_1) = \frac{1}{10}(A+D+E) > 0$$

$$\partial_{1,1}\ell_{log} * \partial_{2,2}\ell_{log} - (\partial_{1,2}\ell_{log})^2$$

$$= \frac{1}{100}((A+B+C+D+E)(A+D+E) - (-A+D+E)(-A+D+E)) > 0$$

Therefore, $\ell_{log}(\beta_0,\beta_1)$ is strictly convex.

The condition for (β_0^*, β_1^*) is :

$$\frac{1}{10}(\frac{3e^{-\beta_1+\beta_0}}{1+e^{-\beta_1+\beta_0}} + \frac{-2e^{-\beta_0}}{1+e^{-\beta_0}} + \frac{e^{\beta_0}}{1+e^{\beta_0}} + \frac{-3e^{-\beta_1-\beta_0}}{1+e^{-\beta_1-\beta_0}} + \frac{e^{\beta_1+\beta_0}}{1+e^{\beta_1+\beta_0}}) = 0$$

and

$$\frac{1}{10}(\frac{-3e^{-\beta_1+\beta_0}}{1+e^{-\beta_1+\beta_0}} + \frac{-3e^{-\beta_1-\beta_0}}{1+e^{-\beta_1-\beta_0}} + \frac{e^{\beta_1+\beta_0}}{1+e^{\beta_1+\beta_0}}) = 0$$

3.

The question $\underset{x\in\mathbb{R}}{max} = \frac{1}{2}\mathbf{x}^T A\mathbf{x}$ subject to $||x||^2 = 1$ is equivalent to the question $\underset{x\in\mathbb{R}}{min} = -\frac{1}{2}\mathbf{x}^T A\mathbf{x}$ subject to $||x||^2 = 1.$

Assume

$$A = \begin{pmatrix} a & b \\ b & d \end{pmatrix}, \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Then

$$f(x_1, x_2) = -\frac{1}{2}\mathbf{x}^T A\mathbf{x} = -\frac{1}{2}(ax_1^2 + 2bx_1x_2 + dx_2^2)$$
$$f'(x_1, x_2) = -\frac{1}{2}\begin{pmatrix} 2a & 2b \\ 2b & 2d \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = -\begin{pmatrix} a & b \\ b & d \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = -A\mathbf{x}$$

$f(x_1, x_2)$ subject to $||x||^2 = 1$, that is $f(x_1, x_2)$ subject to $g(x_1, x_2) = x_1^2 + x_2^2 - 1$

Therefore

$$g'(x_1, x_2) = 2\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Let

$$-\begin{pmatrix} a & b \\ b & d \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 2\lambda\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$-A\mathbf{x} = 2\lambda\mathbf{x}$$

$$A\mathbf{x} = -2\lambda\mathbf{x}$$

Let $\alpha = -2\lambda$, then we have $A\mathbf{x} = \alpha\mathbf{x}$. Therefore α are the eigenvalues of A.

Assume we list all eigenvalues from largest to smallest: $\alpha_1, \alpha_2, \ldots \alpha_n$.

$$-\frac{1}{2}\mathbf{x}^T A\mathbf{x} = -\frac{1}{2}\mathbf{x}^T \alpha\mathbf{x} = -\frac{1}{2}\alpha\mathbf{x}^T\mathbf{x} = -\frac{1}{2}\alpha$$

When $-\frac{1}{2}\mathbf{x}^T A\mathbf{x}$ has the minimum value, $-\frac{1}{2}\alpha$ has the minimum value.

At this time, $\alpha = \alpha_1$, $A\mathbf{x}^* = \alpha_1\mathbf{x}^*$, where $\mathbf{x}^*$ is the corresponding eigenvector of α_1.