DSA3102 HW2
Ethan Keck Jun Wei (April 22, 2024)

Problem 1.

Answer: a) Given,

$$l(\mathbf{w}) = -\frac{1}{n}\sum_{i=1}^{n}\log\sigma(y_i\mathbf{w}^T\mathbf{x}_i), \sigma(\mathbf{z}) = \frac{1}{1+e^{-z}}$$

$$\begin{aligned}\sigma'(\mathbf{z}) &= -\frac{1}{(1+e^{-z})^2}\frac{\partial(1+e^{-z})}{\partial z} \\ &= \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1}{1+e^{-z}}\frac{1+e^{-z}-1}{1+e^{-z}} \\ &= \sigma(z)(1-\sigma(z))\end{aligned}$$

Then the gradient is found as,

$$\begin{aligned}\Rightarrow \nabla l(\mathbf{w}) &= -\frac{1}{n}\sum_{i=1}^{n} y_i\mathbf{x}_i\frac{\sigma(y_i\mathbf{w}^T\mathbf{x}_i)}{\sigma(y_i\mathbf{w}^T\mathbf{x}_i)}(1-\sigma(y_i\mathbf{w}^T\mathbf{x}_i)) \\ &= -\frac{1}{n}\sum_{i=1}^{n} y_i\mathbf{x}_i(1-\sigma(y_i\mathbf{w}^T\mathbf{x}_i))\end{aligned}$$

b) The Hessian is then found as,

$$\begin{aligned}H_l(\mathbf{w}) &= \begin{pmatrix}\partial\nabla l/\partial w_1 & \partial\nabla l/\partial w_2 & \dots & \partial\nabla l/\partial w_p\end{pmatrix} \\ &= \begin{pmatrix}\frac{1}{n}\sum_{i=1}^{n} y_i^2\mathbf{x}_i x_{i,1}\sigma(y_i\mathbf{w}^T\mathbf{x}_i)(1-\sigma(y_i\mathbf{w}^T\mathbf{x}_i)) \\ \frac{1}{n}\sum_{i=1}^{n} y_i^2\mathbf{x}_i x_{i,2}\sigma(y_i\mathbf{w}^T\mathbf{x}_i)(1-\sigma(y_i\mathbf{w}^T\mathbf{x}_i)) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} y_i^2\mathbf{x}_i x_{i,p}\sigma(y_i\mathbf{w}^T\mathbf{x}_i)(1-\sigma(y_i\mathbf{w}^T\mathbf{x}_i))\end{pmatrix}^T \\ &= \frac{1}{n}\sum_{i=1}^{n} y_i^2\sigma(y_i\mathbf{w}^T\mathbf{x}_i)(1-\sigma(y_i\mathbf{w}^T\mathbf{x}_i))\begin{pmatrix}\mathbf{x}_i x_{i,1} & \mathbf{x}_i x_{i,2} & \dots & \mathbf{x}_i x_{i,p}\end{pmatrix} \\ &= \frac{1}{n}\sum_{i=1}^{n} y_i^2\sigma(y_i\mathbf{w}^T\mathbf{x}_i)(1-\sigma(y_i\mathbf{w}^T\mathbf{x}_i))\begin{pmatrix}x_{i,1}x_{i,1} & \dots & x_{i,1}x_{i,p} \\ \vdots & \ddots & \vdots \\ x_{i,p}x_{i,1} & \dots & x_{i,p}x_{i,p}\end{pmatrix} \\ \Rightarrow H_l(\mathbf{w}) &= \frac{1}{n}\sum_{i=1}^{n} y_i^2\mathbf{x}_i\mathbf{x}_i^T\sigma(y_i\mathbf{w}^T\mathbf{x}_i)(1-\sigma(y_i\mathbf{w}^T\mathbf{x}_i))\end{aligned}$$

Let $l_i(\mathbf{w}) = \log\sigma(y_i\mathbf{w}^T\mathbf{x}_i)$ be such that

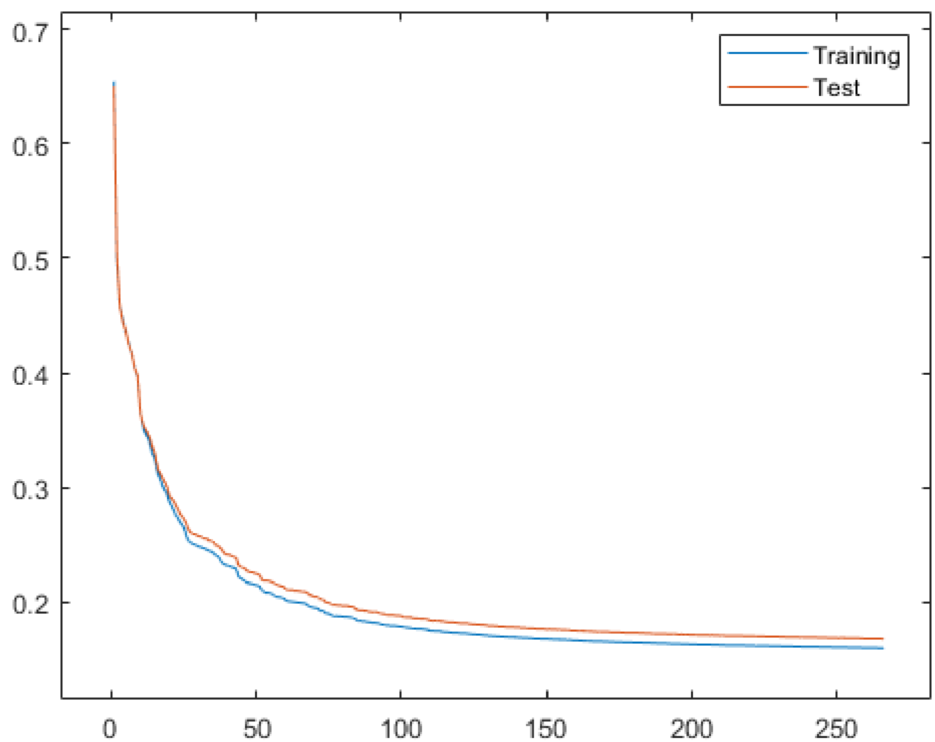
$$l(\mathbf{w}) = -\frac{1}{n}\sum_{i=1}^{n} l_i(\mathbf{w})$$

$$\Rightarrow H_{l_i}(\mathbf{w}) = y_i^2\mathbf{x}_i\mathbf{x}_i^T\sigma(y_i\mathbf{w}^T\mathbf{x}_i)(1-\sigma(y_i\mathbf{w}^T\mathbf{x}_i))$$

Since $\sigma(\mathbf{z}) \in [0,1]$ for all $\mathbf{z}$, $\Rightarrow 1 - \sigma(\mathbf{a}) \in [0,1]$,
then $y_i^2 \sigma(y_i \mathbf{w}^T \mathbf{x}_i)(1 - \sigma(y_i \mathbf{w}^T \mathbf{x}_i)) \geq 0$ for all $y_i \mathbf{w}^T \mathbf{x}_i$.
We check if $\mathbf{x}_i \mathbf{x}_i^T$ is positive definite;
For all $\mathbf{v} \in R^p$,

$$\mathbf{v}^T \mathbf{x}_i \mathbf{x}_i^T \mathbf{v} = (\mathbf{x}_i^T \mathbf{v})^T (\mathbf{x}_i^T \mathbf{v}) = (\mathbf{x}_i^T \mathbf{v})^2 \geq 0$$
$$\Rightarrow y_i^2 \sigma(y_i \mathbf{w}^T \mathbf{x}_i)(1 - \sigma(y_i \mathbf{w}^T \mathbf{x}_i)) \cdot (\mathbf{x}_i^T \mathbf{v})^2 \geq 0$$

Hence $H_{l_i}(\mathbf{w}) = y_i^2 \mathbf{x}_i \mathbf{x}_i^T \sigma(y_i \mathbf{w}^T \mathbf{x}_i)(1 - \sigma(y_i \mathbf{w}^T \mathbf{x}_i))$ is positive definite and $l_i(\mathbf{w})$ is convex, and by Corollary 3.1, $l(\mathbf{w})$ is therefore convex.
c) & d)



The best best model parameter is

$$\mathbf{w}^* = (-0.0875, -0.0906, -0.0897, 0.2244, 0.4252, 0.2253, 0.8877, 0.4803, 0.1037, 0.1133, 0.0337,$$
$$-0.1983, -0.1725, 0.1954, 0.2317, 0.5274, 0.4339, 0.0013, 0.0671, 0.2169, 0.2119,$$
$$0.1843, 0.5949, 0.5848, -0.8879, -0.2271, -0.9779, 0.0354, -0.1465, -0.0386, -0.1091,$$
$$-0.0029, -0.1703, -0.0106, -0.1612, 0.2882, -0.4286, 0.1027, -0.2027, -0.0366, -0.1122,$$
$$-0.5082, -0.1578, -0.3256, -0.3231, -0.7341, 0.0547, -0.2213, -0.1272, -0.0401, -0.0062,$$
$$0.7794, 0.9977, 0.1117, 0.4161, 0.2900, 0.1242)$$

with a loss value of $l(\mathbf{w}^*) = 0.160133415155724$ for the training data
and $l(\mathbf{w}^*) = 0.168318657862923$ for the test data. $\square$

Problem 2.

Answer: a)

$$\min f(\mathbf{x}) := x_1 - x_2$$
$$g_1(\mathbf{x}) := x_1^2 + x_2^2 - 2 = 0$$
$$g_2(\mathbf{x}) := x_2 - x_3^2 = 0$$

Given $g_1(\mathbf{x}) := x_1^2 + x_2^2 = 2$ is a circle
on the x_1x_2 plane centered on (0, 0) with $radius = \sqrt{2}$ then

$$-\sqrt{2} \leq x_1 \leq \sqrt{2}, -\sqrt{2} \leq x_2 \leq \sqrt{2},$$

Since

$$g_2(\mathbf{x}) := x_2 = x_3^2, -\sqrt{2} \leq x_2 \leq \sqrt{2}$$
$$\Rightarrow -\sqrt{2} \leq x_3^3 \leq \sqrt{2}$$
$$\Rightarrow -2^{1/6} \leq x_3 \leq 2^{1/6}$$
$$\Rightarrow -2^{1/6} \leq x_i \leq 2^{1/6}, \forall i = 1, 2, 3$$

Hence $\alpha = 2^{1/6}$
b) We show that $\nabla g_1(\mathbf{x})$ and $\nabla g_2(\mathbf{x})$ are linearly independent.

$$\begin{pmatrix} \nabla g_1(\mathbf{x}) & \nabla g_2(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} 2x_1 & 0 \\ 2x_2 & 1 \\ 0 & -3x_3^2 \end{pmatrix}$$

We prove the gradients are linearly independent by contradiction; i.e. prove they cannot be parallel by contradiction. Let

$$\begin{pmatrix} 2x_1 \\ 2x_2 \\ 0 \end{pmatrix} = \lambda \begin{pmatrix} 0 \\ 1 \\ -3x_3^2 \end{pmatrix}$$
$$\Rightarrow x_1 = 0, x_2 = \lambda/2$$
$$\Rightarrow g_1(\mathbf{x}) = 0$$
$$\text{From third components, } \lambda^2 = 8 \Rightarrow \lambda \neq 0, x_3 = 0$$
$$\text{but then } g_2(\mathbf{x}) = 0 \Rightarrow x_2 = 0 \text{ which contradicts } x_2 = \lambda/2 \neq 0$$

c) For $x \in S$, the KKT conditions are;

$$\nabla f(\mathbf{x}) + \lambda_1 \nabla g_1(\mathbf{x}) + \lambda_2 \nabla g_2(\mathbf{x})$$
$$= \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} + \lambda_1 \begin{pmatrix} 2x_1 \\ 2x_2 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ 1 \\ -3x_3^2 \end{pmatrix} = \mathbf{0}$$
$$\Rightarrow \begin{cases} 1 + 2\lambda_1 x_1 = 0 \\ -1 + 2\lambda_1 x_2 + \lambda_2 = 0 \\ -3\lambda_2 x_3^2 = 0 \end{cases}$$

From the third eqn, $x_3 = 0$ or $\lambda_2 = 0$,

If $x_3 = 0$, $g_2(\mathbf{x}) = 0 \Rightarrow x_2 = 0$, then from the second equation $\lambda_2 = 1$.
Then $g_1(\mathbf{x}) = 0, x_1 = \pm\sqrt{2}$ and the first equation gives $\lambda_1 = \mp 1/(2\sqrt{2})$.
Hence, $\mathbf{x}^{(1)} = \begin{pmatrix} -\sqrt{2} & 0 & 0 \end{pmatrix}^T$ is a KKT solution, with $f(\mathbf{x}^{(1)}) = -\sqrt{2}$ and
$\mathbf{x}^{(2)} = \begin{pmatrix} \sqrt{2} & 0 & 0 \end{pmatrix}^T$ is a KKT solution, with $f(\mathbf{x}^{(2)}) = \sqrt{2}$.
If $\lambda_2 = 0$, then when rearranging, we have

$$x_1 = -\frac{1}{2\lambda_1}, x_2 = \frac{1}{2\lambda_1},$$

Then for $g_1(\mathbf{x}) = 0 \Rightarrow \lambda_1 = \pm 1/2$
If $\lambda_1 = 1/2$, then $x_1 = -1, x_2 = 1$ and $g_2(\mathbf{x}) = 0 \Rightarrow x_3 = 1$.
If $\lambda_1 = -1/2$, then $x_1 = 1, x_2 = -1$ and $g_2(\mathbf{x}) = 0 \Rightarrow x_3 = -1$.
$\mathbf{x}^{(3)} = \begin{pmatrix} -1 & 1 & 1 \end{pmatrix}^T$ is a KKT solution, with $f(\mathbf{x}^{(3)}) = -2$ and
$\mathbf{x}^{(4)} = \begin{pmatrix} 1 & -1 & -1 \end{pmatrix}^T$ is a KKT solution, with $f(\mathbf{x}^{(4)}) = 2$.
The global minimizer is thus $\mathbf{x}^{(3)}$. $\square$

Problem 3.

Answer: a) Let

$$\begin{gathered}
h_1(x) := x^2 > 0, x \in (-\infty, \infty) \setminus \{0\} \\
h_2(x) := x + 1 > 0, \\
\Rightarrow x > -1, x \in (-1, \infty) \\
\Rightarrow F^< = (-1, \infty) \setminus \{0\}
\end{gathered}$$

b) Given

$$\begin{gathered}
P(x, \mu) = x - \mu \log(x^2) + \mu \log(x+1) \\
\frac{\partial P(x,\mu)}{\partial x} = 1 - \mu \left(\frac{2}{x} + \frac{1}{x+1} \right) = 0 \\
\Rightarrow \frac{2x + 2 + x}{x(x+1)} = \frac{1}{\mu} \\
\Rightarrow \mu(3x + 2) = x(x+1) \\
\Rightarrow x^2 + x - 3\mu x - 2\mu = 0 \\
\Rightarrow x^2 + (1 - 3\mu)x - 2\mu = 0 \\
x = \frac{(3\mu - 1) \pm \sqrt{(1-3\mu)^2 - 4(-2\mu)}}{2} \\
= \frac{(3\mu - 1) \pm \sqrt{(1-3\mu)^2 + 8\mu}}{2} \\
\Rightarrow x_\mu \in \left\{ x = \frac{(3\mu - 1) \pm \sqrt{(1-3\mu)^2 + 8\mu}}{2} \;\middle|\; \mu > 0 \right\}
\end{gathered}$$

$$\begin{gathered}
\lim_{\mu \to 0} x_\mu = \lim_{\mu \to 0} \frac{(3\mu - 1) \pm \sqrt{(1-3\mu)^2 + 8\mu}}{2} \\
= \lim_{\mu \to 0} \frac{(3(0) - 1) \pm \sqrt{(1-3(0))^2 + 8(0)}}{2} \\
= \frac{-1 \pm \sqrt{1}}{2} \\
\Rightarrow \lim_{\mu \to 0} x_\mu = -1, \lim_{\mu \to 0} x_\mu = 0
\end{gathered}$$

$\square$

Problem 4.

Answer: a) NLP: Let $V(x,y,z)$ be the volume enclosed by the box,

$$\begin{gathered}
\min_{x>0,y>0,z>0} -V(x,y,z) = -xyz \\
4(x+y+z) - 20 = 0 \text{ (length of all edges)} \\
\Rightarrow g_1(\mathbf{x}) := x+y+z-5 = 0 \\
2(xy+xz+yz) - 16 = 0 \text{ (area of all faces)} \\
\Rightarrow g_2(\mathbf{x}) := xy+xz+yz-8 = 0
\end{gathered}$$

b) To find the feasible set

$$\begin{gathered}
x \cdot g_1(\mathbf{x}) := x^2+xy+xz = 5x \\
y \cdot g_1(\mathbf{x}) := xy+y^2+yz = 5y \\
z \cdot g_1(\mathbf{x}) := xz+yz+z^2 = 5z \\
\Rightarrow x \cdot g_1(\mathbf{x}) + y \cdot g_1(\mathbf{x}) + z \cdot g_1(\mathbf{x}) \\
\Rightarrow x^2+y^2+z^2+2(xy+yz+xz) = 5(x+y+z) \\
\Rightarrow x^2+y^2+z^2+16 = 25 \\
\Rightarrow x^2+y^2+z^2 = 3^2
\end{gathered}$$

$x^2+y^2+z^2=3^2$ is a sphere centered at (0,0,0) with a radius of 3.
Hence, the feasible set is $F = \{(x,y,z) | 0 < x,y,z < 3, x^2+y^2+z^2 = 3^2\}$.
Since the box is not a cube, $x \neq y \neq z$. Given the KKT conditions;

$$\begin{gathered}
L(\mathbf{x},\lambda) = -xyz + \lambda_1(x+y+z-5) + \lambda_2(xy+xz+yz-8) \\
-\nabla V(\mathbf{x}) + \lambda_1 \nabla g_1(\mathbf{x}) + \lambda_2 \nabla g_2(\mathbf{x}) = \mathbf{0} \\
\Rightarrow - \begin{pmatrix} yz \\ xz \\ xy \end{pmatrix} + \lambda_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} y+z \\ x+z \\ x+y \end{pmatrix} = \mathbf{0} \\
\Rightarrow \begin{cases} -yz + \lambda_1 + \lambda_2(y+z) = 0 - (1) \\ -xz + \lambda_1 + \lambda_2(x+z) = 0 - (2) \\ -xy + \lambda_1 + \lambda_2(x+y) = 0 - (3) \\ x+y+z-5 = 0 - (4) \\ xy+xz+yz-8 = 0 - (5) \end{cases}
\end{gathered}$$

We check that $\nabla g_1(\mathbf{x})$ and $\nabla g_1(\mathbf{x})$ are linearly independent.

$$\left(\begin{array}{cc|c} 1 & y+z & 0 \\ 1 & x+z & 0 \\ 1 & x+y & 0 \end{array} \right) \xrightarrow{\text{Gaussian Elimination}} \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{array} \right)$$

Hence, there exists only the unique solution of $\mathbf{0}$, and thus $\nabla g_1(\mathbf{x})$, $\nabla g_2(\mathbf{x})$ are linearly independent, and all $\mathbf{x} > \mathbf{0}$ are regular points. Then,

$$\begin{gathered} \Rightarrow (1)-(2) \Rightarrow z(x-y) - \lambda_2(x-y) = 0, z = \lambda_2 \\ \Rightarrow (1)-(3) \Rightarrow y(x-z) - \lambda_2(x-z) = 0, y = \lambda_2 \\ \Rightarrow x+y+z = 5, x + 2\lambda_2 = 5, x = 5 - 2\lambda_2 \end{gathered}$$

$$\begin{gathered} \theta(\lambda) = L(x = 5-2\lambda_2, y = \lambda_2, z = \lambda_2, \lambda) \\ = -(5-2\lambda_2)(\lambda_2)^2 + \lambda_1(5 - 2\lambda_2 + \lambda_2 + \lambda_2 - 5) \\ +\lambda_2((5-2\lambda_2)\lambda_2 + (5-2\lambda_2)\lambda_2 + \lambda_2^2 - 8) \end{gathered}$$

$$\begin{gathered} = -5\lambda_2^2 + 2\lambda_2^3 + \lambda_2^3 - 8\lambda_2 + 2\lambda_2^2(5 - 2\lambda_2) \\ = 3\lambda_2^3 - 4\lambda_2^3 - 5\lambda_2^2 + 10\lambda_2^2 - 8\lambda_2 \\ = -\lambda_2^3 + 5\lambda_2^2 - 8\lambda_2, \lambda_2 \in R \\ \Rightarrow \theta'(\lambda_2) = -3\lambda_2^2 + 10\lambda_2 - 8 = 0 \\ \lambda_2 = \frac{4}{3}, \lambda_2 = 2 \end{gathered}$$

Then

$$\begin{cases} x = 5 - 2\left(\frac{4}{3}\right) = \frac{7}{3}, x = 5 - 2(2) = 1 \\ y = \frac{4}{3}, y = 2 \\ z = \frac{4}{3}, z = 2 \end{cases}$$

respectively.
Repeating the above for other combinations with (4),

$$\begin{gathered} (2)-(3) \Rightarrow x(y-z) - \lambda_2(y-z) = 0, x = \lambda_2 \\ \Rightarrow \begin{cases} y = 5 - 2\lambda_2, \text{ if } x = \lambda_2, z = \lambda_2 \\ z = 5 - 2\lambda_2, \text{ if } x = \lambda_2, y = \lambda_2 \end{cases} \end{gathered}$$

$$\Rightarrow \begin{cases} x = \frac{4}{3}, x = 2 \\ y = 5 - 2\left(\frac{4}{3}\right) = \frac{7}{3}, y = 5 - 2(2) = 1 \\ z = \frac{4}{3}, z = 2 \end{cases}, \begin{cases} x = \frac{4}{3}, x = 2 \\ y = \frac{4}{3}, y = 2 \\ z = 5 - 2\left(\frac{4}{3}\right) = \frac{7}{3}, z = 5 - 2(2) = 1 \end{cases}$$

respectively.
We have KKT points as
$\mathbf{x}^* \in \{\frac{1}{3}(7,4,4), \frac{1}{3}(4,7,4), \frac{1}{3}(4,4,7), (1,2,2), (2,1,2), (2,2,1)\}$, we verify feasibility,

$$\begin{gathered} (7/3)^2 + (4/3)^2 + (4/3)^2 = 9 \text{ (Feasible)}, \\ (1)^2 + (2)^2 + (2)^2 = 9 \text{ (Feasible)} \end{gathered}$$

and duality gap,

$$\begin{aligned}
-V(7/3,4/3,4/3) &= -V(4/3,7/3,4/3) = -V(4/3,4/3,7/3) \\
-\frac{7}{3}\cdot\frac{4}{3}\cdot\frac{4}{3} &= -\frac{112}{27} \approx -4.1481 \\
&= \theta(\lambda_2 = 4/3) = -\left(\frac{4}{3}\right)^3 + 5\left(\frac{4}{3}\right)^2 - 8\left(\frac{4}{3}\right)
\end{aligned}$$

$$\begin{aligned}
-V(1,2,2) &= -V(2,1,2) = -V(2,2,1) = -1\cdot 2\cdot 2 = -4 \\
&= \theta(\lambda_2 = 2) = -(2)^3 + 5(2)^2 - 8(2)
\end{aligned}$$

Hence the optimal global minimizers for $-V(\mathbf{x})$ are $(7/3,4/3,4/3)$, $(4/3,7/3,4/3)$, $(4/3,4/3,7/3)$ $\square$