<u>Home Assignment 3</u>

Daniel Smotritsky, Ilana Pervoi

1. Convexity

(a)

i. **e^{ax} is convex.**

Proof:
$a, x \in \mathbb{R}$ and as we have learned, $\mathbb{R}$ is a convex region.

$$f'(x) = ae^{ax}$$
$$f''(x) = a^2e^{ax}$$
$$f''(x) \geq 0 \; as \; a^2 \geq 0, \qquad e^{ax} \geq 0$$

→ From definition 3 to convexity, e^{ax} is convex. ■

ii. **$-\log(x)$ is concave when the log base is in [0,1) and convex when the log base >1.**

Proof:
$x \in \mathbb{R}^+$ is a convex region:

Let $\alpha \in [0,1]$ and $x, y \in \mathbb{R}^+$
Let us look at $\alpha x + (1-\alpha)y$:
<u>For $\alpha = 0$:</u> $0 * x + (1-0) * y = y \in \mathbb{R}^+$ from assumption.
<u>For $\alpha = 1$:</u> $1 * x + (1-1) * y = x \in \mathbb{R}^+$ from assumption.
<u>For $\alpha \in (0,1)$:</u> $\alpha, 1-\alpha \in \mathbb{R}^+$

$$\rightarrow \; \alpha x, (1-\alpha)y \in \mathbb{R}^+$$
$$\rightarrow \; \alpha x + (1-\alpha)y \in \mathbb{R}^+$$

Hence from the definition of convex set, $\mathbb{R}^+$ is a convex region. ■

Assume a is the log base:

$$f'(x) = -\frac{1}{xln(a)} = -\frac{1}{x} * \frac{1}{\ln(a)}$$
$$f''(x) = \frac{1}{x^2} * \frac{1}{\ln(a)}$$
$$\frac{1}{x^2} > 0$$

$$\begin{cases} if\ 0 \le a < 1 \to \ln(a) < 0 \to f''(x) < 0 \to f\ is\ concave \\ if\ a = 1 \to \ln(a) = 0 \to f''(x)\ is\ undefined \\ if\ a > 1 \to \ln(a) > 0 \to f''(x) > 0 \to f\ is\ convex \end{cases}$$

From definition 3 to convexity. ■

iii. **log (x) is convex when the log base is in [0,1) and concave when the log base >1.**

Proof:
$x \in \mathbb{R}^+$ is a convex region as proved in ii.

Assume a is the log base:

$$f'(x) = \frac{1}{xln(a)} = \frac{1}{x} * \frac{1}{\ln(a)}$$

$$f''(x) = -\frac{1}{x^2} * \frac{1}{\ln(a)}$$

$$-\frac{1}{x^2} < 0$$

$$\begin{cases} if\ 0 \le a < 1 \to \ln(a) < 0 \to f''(x) > 0 \to f\ is\ convex \\ if\ a = 1 \to \ln(a) = 0 \to f''(x)\ is\ undefined \\ if\ a > 1 \to \ln(a) > 0 \to f''(x) < 0 \to f\ is\ concave \end{cases}$$

From definition 3 to convexity. ■

iv. **$|x|^a, a \ge 1$ is convex.**

Proof:
$x \in \mathbb{R}$ is a convex region.

- <u>If a=1:</u>

$$f(x) = |x|$$

Let $\alpha \in [0,1]$ and $x, y \in \mathbb{R}$.

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha)f(y) \leftrightarrow$$

$$|\alpha x + (1-\alpha)y| \le \alpha|x| + (1-\alpha)|y| \leftrightarrow$$

$$\alpha, 1-\alpha \ge 0$$

$$|\alpha x + (1-\alpha)y| \le |\alpha x| + |(1-\alpha)y|$$

$$true\ from\ the\ triangle\ inequality.$$

→ From definition 1 to convexity, $|x|$ is convex. ■

- <u>If a>1:</u>

$$f(x) = |x|^a = \begin{cases} (-x)^a, & x < 0 \\ x^a, & x \geq 0 \end{cases}$$

$|x|^a$ is differentiable:

<u>For $x \neq 0$</u>: $|x|^a$ is differentiable as $x^a, (-x)^a$ are sleek and monotone.

<u>For $x = 0$:</u>

$$f'(0) = 0$$

We can observe by the definition of differentiability:

$$f'(0) = \lim_{h \to 0} \frac{f(0+h) - f(0)}{h} = \lim_{h \to 0} \frac{f(h) - 0}{h} = \lim_{h \to 0} \frac{|h|^a}{h}$$
$$= \begin{cases} \lim\limits_{h \to 0^+} h^{a-1}, & h \geq 0 \\ \lim\limits_{h \to 0^-} (-h)^{a-1}, & h < 0 \end{cases} =$$
$$= \begin{cases} 0, & h \geq 0 \\ 0, & h < 0 \end{cases}$$

Meaning,

$$0 = f'(0) = \lim_{h \to 0} \frac{f(0+h) - f(0)}{h} = 0$$

$\rightarrow$ $|x|^a$ is differentiable for all x, and:

$$f'^{(x)} = \begin{cases} -a * (-x)^{a-1}, & x < 0 \\ a * x^{a-1}, & x \geq 0 \end{cases}$$

Let $x, y \in \mathbb{R}$.

$$f(x) \geq f(y) + f'(y) * (x - y) \leftrightarrow$$
$$g(x) = f(x) - f(y) + f'(y) * (x - y) \geq 0$$

We can observe that when $x = y$, $g(y) = 0$:

$$g(y) = f(y) - f(y) - f'(y) * (y - y) = 0 - f'(y) * 0 = 0$$

Let us look at $g'(x)$ around $x = y$:

$$g'(x) = f'(x) - f'(y)$$
$$= \begin{cases} -a * (-x)^{a-1} - (-a * (-y)^{a-1}) = -a((-x)^{a-1} - y^{a-1}), & x < 0 \ and \ y < 0 \ (*) \\ -a * (-x)^{a-1} - (a * y^{a-1}) = -a((-x)^{a-1} + y^{a-1}), & x < 0 \ and \ y \geq 0 \ (**) \\ a * x^{a-1} - (-a * (-y)^{a-1}) = a(x^{a-1} + (-y)^{a-1}), & x \geq 0 \ and \ y < 0 \ (***) \\ a * x^{a-1} - a * y^{a-1} = a(x^{a-1} + y^{a-1}), & x \geq 0 \ and \ y \geq 0 \ (****) \end{cases}$$

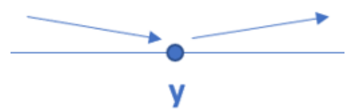
$\underline{if\ x \leq y < 0}$: From (*) $g'(x) = -a((-x)^{a-1} - y^{a-1}) > 0$

$\underline{if\ x < 0 \leq y}$: From (**) $g'(x) = -a((-x)^{a-1} + y^{a-1}) < 0$

$\underline{if\ y < 0 \leq x}$: From (***) $g'(x) = a(x^{a-1} + (-y)^{a-1}) < 0$

$\underline{if\ y \leq x < 0}$: From (****) $g'(x) = a(x^{a-1} + y^{a-1}), > 0$

Overall, for $x \geq y$, $g'(x) > 0$ and for $x < y, g'(x) < 0$



Meaning, x=y is global minimum, $g(y) = 0$, hence $g(x) \geq 0$ for all $x \in \mathbb{R}$.

→ From definition 2 to convexity, $|x|^a$ is convex. ■

v. **x^3 is not convex and not concave.**

Proof:
$x \in \mathbb{R}$ is a convex region.

$$f'(x) = 3x^2$$
$$f''(x) = 6x$$

$$\rightarrow \begin{cases} f''(x) > 0, x > 0 \\ f''(x) \leq 0, x \leq 0 \end{cases}$$

→ From definition 3 to convexity, x^3 is not convex and not concave. ■

(b) **$f(x) = x^T Ax + b^T x + c$ is convex under the condition that A is PSD.**

Let $\alpha \in [0,1]$ and $x, y \in \mathbb{R}^n$.
We will explore when definition 1 to convexity is obtained:

$$\boldsymbol{f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y)} \leftrightarrow$$

$$
\begin{aligned}
&(\alpha x + (1-\alpha)y)^T A(\alpha x + (1-\alpha)y) + c \\
&\qquad \le \alpha(x^T Ax + b^T x + c) + (1-\alpha)(y^T Ay + b^T y + c) \leftrightarrow
\end{aligned}
$$

$$
\begin{aligned}
&(\alpha x^T A + (1-\alpha)y^T A)(\alpha x + (1-\alpha)y) + b^T(\alpha x + (1-\alpha)y) + \cancel{c} \\
&\qquad \le \alpha x^T Ax + \alpha b^T x + \cancel{\alpha c} + y^T Ay + b^T y + \cancel{c} - \alpha y^T Ay - \alpha b^T y \\
&\qquad \cancel{-\alpha c} \leftrightarrow
\end{aligned}
$$

$$
\begin{aligned}
&\alpha x^T Ax + (1-\alpha)\alpha x^T Ay + (1-\alpha)\alpha y^T Ax + (1-\alpha)^2 y^T Ay + \cancel{\alpha b^T x} \\
&\qquad + \cancel{(1-\alpha) b^T y} \\
&\qquad \le \alpha x^T Ax + \cancel{\alpha b^T x} + y^T Ay + \cancel{b^T y} - \alpha y^T Ay \cancel{- \alpha b^T y} \leftrightarrow
\end{aligned}
$$

$$
\begin{aligned}
&\alpha x^T Ax + \alpha x^T Ay - \alpha^2 x^T Ay + \alpha y^T Ax - \alpha^2 y^T Ax + (1 - 2\alpha + \alpha^2) y^T Ay \\
&\qquad \le \alpha x^T Ax + y^T Ay - \alpha y^T Ay \leftrightarrow
\end{aligned}
$$

$$-\alpha(1-\alpha)[x^T Ax - x^T Ay - y^T Ax + y^T Ay] \le 0 \leftrightarrow$$

$$\alpha(1-\alpha)[(x-y)^T A(x-y)] \ge 0 \leftrightarrow$$

$$\alpha, 1-\alpha \ge 0$$

$$(x-y)^T A(x-y) \ge 0 \leftrightarrow$$

$$Denote\ z = x - y$$

$$z^T Az \ge 0 \leftrightarrow \boldsymbol{A\ is\ PSD}$$

(c) Let $f: \mathbb{R}^n \to \mathbb{R}$ differentiable in a convex domain Ω.
Need to prove:

$$f\ is\ convex\ \leftrightarrow f(y) \ge f(x) + \nabla f(x)^T(y-x), \qquad \forall x, y \in \Omega$$

($\to$): Assume f is convex.
Let $\theta \in [0,1]$ and $x, y \in \mathbb{R}^n$.
From convexity we obtain:

$$f(\theta y + (1-\theta)x) \le \theta f(y) + (1-\theta) f(x)$$

$$
\begin{aligned}
f(\theta y + (1-\theta)x) &= f(\theta y + x - \theta x) = f\big(x + \theta(y-x)\big) \\
\theta f(y) + (1-\theta) f(x) &= \theta f(y) + f(x) - \theta f(x) = f(x) + \theta(f(x) - f(y)) \\
\to\ f\big(x + \theta(y-x)\big) &\le f(x) + \theta\big(f(x) - f(y)\big) \quad (*)
\end{aligned}
$$

Now, from Taylor expansion, when $\theta \to 0$:

$$f\big(x+\theta(y-x)\big) = f(x) + f'(x) * \theta(y-x)$$
$$= f(x) + +\theta\nabla f(x)^T(y-x) \ \ (**)$$

From (*) and (**):

$$\cancel{f(x)} + \cancel{\theta}\nabla f(x)^T(y-x) \ = f\big(x+\theta(y-x)\big) \le \cancel{f(x)} + \cancel{\theta}\big(f(x)-f(y)\big)$$
$$\rightarrow \ \nabla f(x)^T(y-x) \le f(x) - f(y)$$
$$\rightarrow f(y) \ge f(x) + \nabla f(x)^T(y-x) \ \blacksquare$$

(←): Assume $f(y) \ge f(x) + \nabla f(x)^T(y-x), \ \forall x,y \in \ \Omega$

Let $t = (1-\theta)x + \theta y$ be a point between x, y. $\theta \in (0,1)$.

By definition 1 for convexity, we need to prove:

$$f(t) = f((1-\theta)x + \theta y) \le \theta f(y) + (1-\theta)f(x)$$

From the assumption:

$$\begin{cases} f(y) \ge f(t) + \nabla f(t)^T(y-t) \ \ /* \ \theta \\ f(x) \ge f(t) + \nabla f(t)^T(x-t) \ \ /* \ (1-\theta) \end{cases}$$

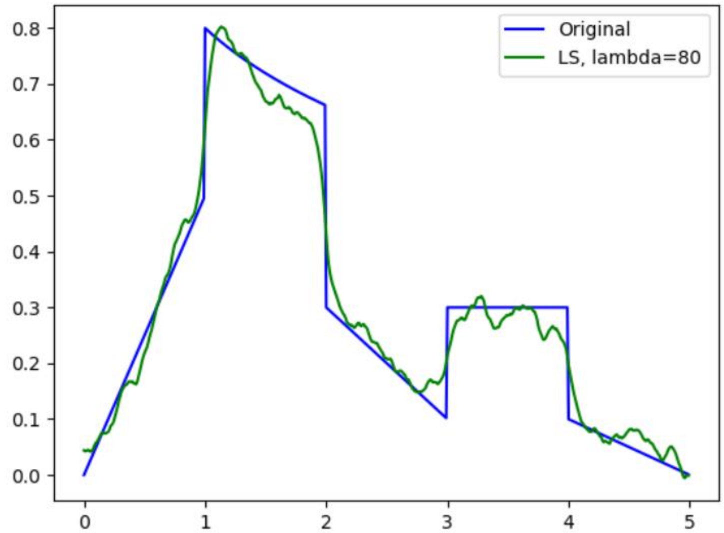
$$\rightarrow \begin{cases} \theta\, f(y) \ge \theta f(t) + \theta\nabla f(t)^T(y-t) \\ (1-\theta)f(x) \ge (1-\theta)f(t) + (1-\theta)\nabla f(t)^T(x-t) \end{cases}$$

By summing these two equations we get:

$$\begin{aligned} \theta\, f(y) + (1-\theta)f(x) & \\ &\ge \cancel{\theta f(t)} + \theta\nabla f(t)^T(y-t) + (1\cancel{-\theta})f(t) \\ &+ (1-\theta)\nabla f(t)^T(x-t) \\ &= f(t) + \theta\nabla f(t)^T[\theta(y-t) + (1-\theta)(x-t)] \\ &= \ f(t) + \theta\nabla f(t)^T[\theta y + (1-\theta)x - \theta t + (1-\theta)t] \\ &= \ f(t) + \theta\nabla f(t)^T[\cancel{t - \theta t + (1-\theta)t}] = f(t) + \cancel{\theta\nabla f(t)^T * 0} \\ &= f(t) \ \blacksquare \end{aligned}$$

2. Iterative re-weighted least squares (IRLS) for 1D Total Variation

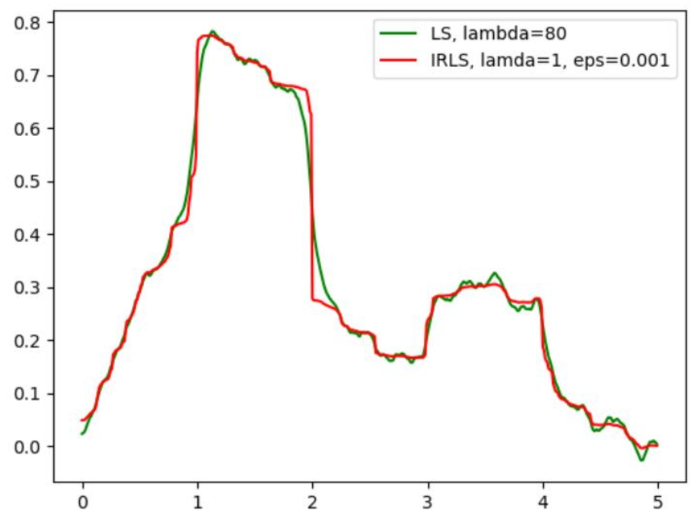
(a)

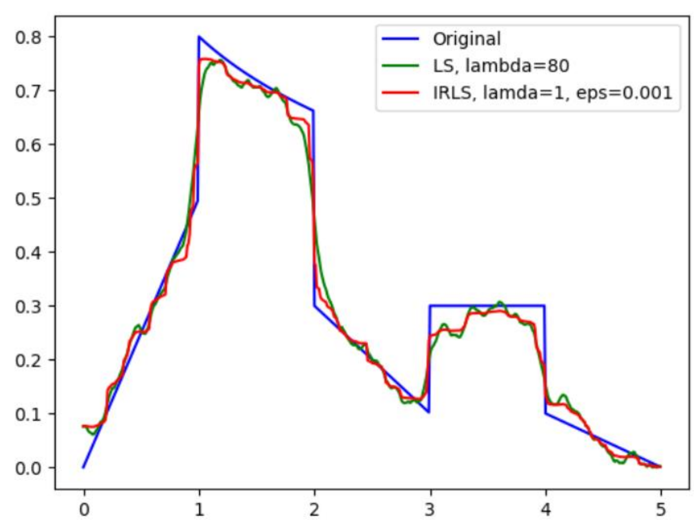



```
x = np.arange(0, 5, 0.01)
n = np.size(x)
one = int(n / 5)
f = np.zeros(x.shape)
f[0:one] = 0.0 + 0.5 * x[0:one]
f[one:2 * one] = 0.8 - 0.2 * np.log(x[100:200])
f[(2 * one):3 * one] = 0.7 - 0.2 * x[(2 * one):3 * one]
f[(3 * one):4 * one] = 0.3
f[(4 * one):(5 * one)] = 0.5 - 0.1 * x[(4 * one):(5 * one)]
G = spdiags([-np.ones(n), np.ones(n)], np.array([0, 1]), n,
n).toarray()
A = np.eye(n)
etta = 0.1 * np.random.randn(np.size(x))
y = f + etta

# regularized non-weighted LS L2
AT = np.transpose(A)
lamb = 80
fn = np.linalg.inv(AT @ A + (lamb / 2) * np.transpose(G) @ G)
@ AT @ y
plt.figure()
plt.plot(x, f, 'b', label="Original")
plt.plot(x, fn, 'g',  label="LS, lambda=80")
plt.legend()
plt.show()
```

(b)

```
# IRLS
W = np.eye(n)
GT = np.transpose(G)
lamb = 1
eps = 0.001

for k in range(10):
    x_k = np.linalg.inv(AT @ A + (lamb / 2) * GT @ W @ G) @
AT @ y
    Gx_k = G @ x_k
    W_diag = [1 / (np.abs(Gx_k[i]) + eps) for i in range(n)]
    W = np.diag(W_diag)
```



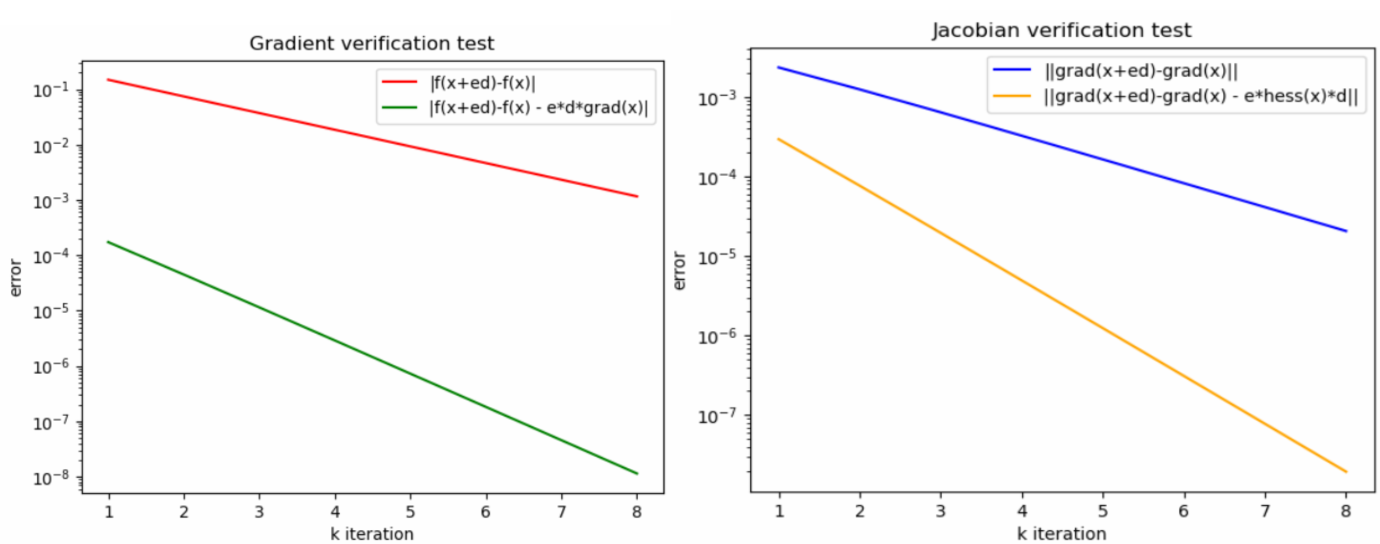


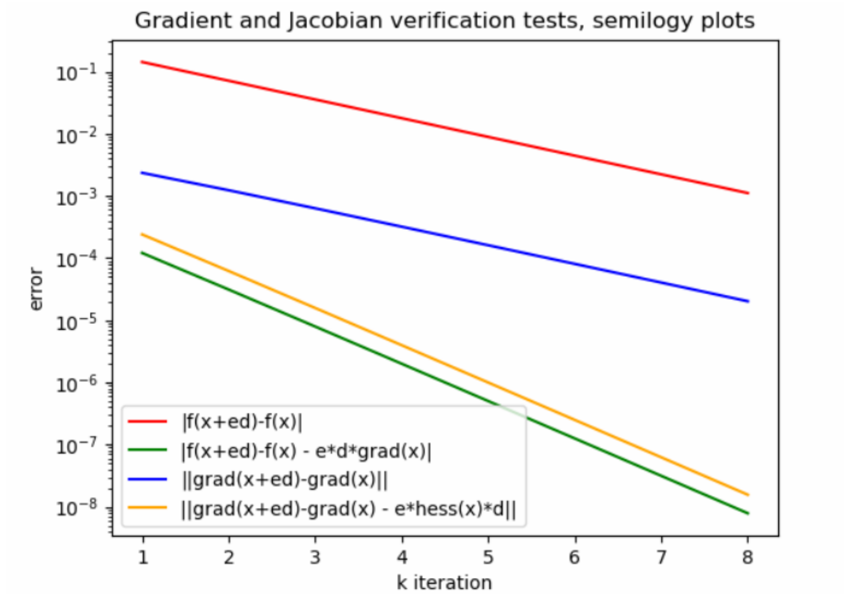
4. Non-linear convex optimization for handwritten digits classification

(a)

```
def sigmoid(x):
    return 1.0 / (1 + np.exp(-x))


def func(X, w, labels, h=False):
    # Linear Regression Objective
    c1 = labels
    c2 = 1 - c1
    m = np.shape(X)[1]
    sigm_XTw = sigmoid(np.transpose(X) @ w)
    Fw = (-1 / m) * (np.transpose(c1) @ np.log(sigm_XTw)
                     + np.transpose(c2) @ np.log(1 - sigm_XTw))

    # Gradient
    Grad = (1 / m) * X @ (sigm_XTw - c1)

    if h:
        # Hessian
        D_diag = np.multiply(sigm_XTw, 1 - sigm_XTw)
        D = np.diag(D_diag)
        Hess = (1 / m) * X @ D @ np.transpose(X)
        return Fw, Grad, Hess

return Fw, Grad
```

(b)



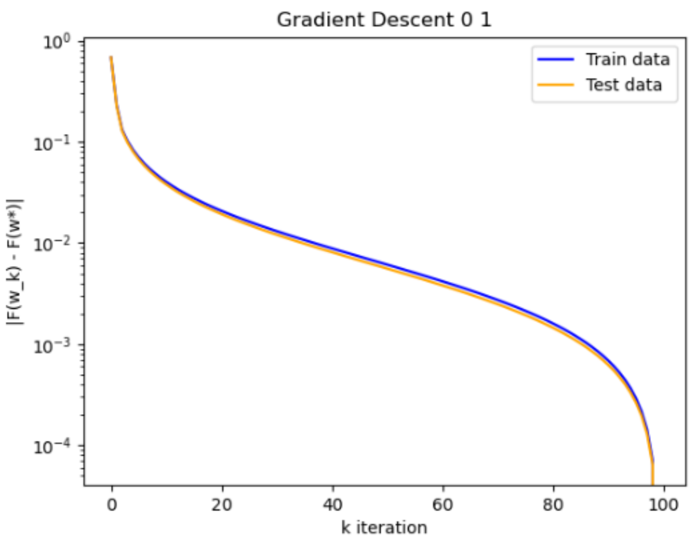


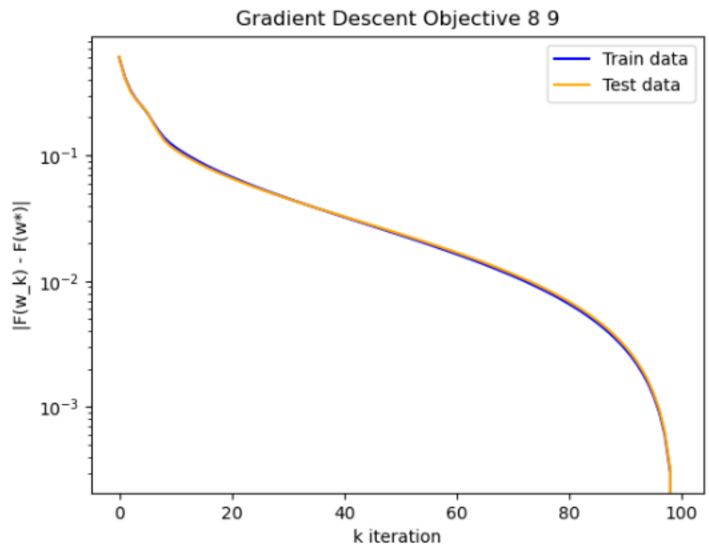
(c)

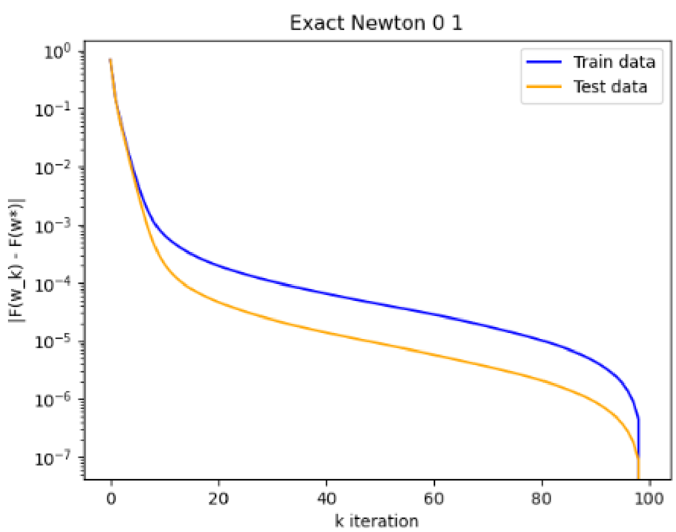
```
def gradient_descent(A, x, labels):
    c1 = np.array(labels)
    x_axis = []
    y_axis = []

    for i in range(100):
        Fx, grad_x = func(A, x, c1)
        d = -1 * grad_x
        alpha = linesearch(x, Fx, grad_x, d, 0.25, 0.5, 1e-4, c1, A)

        x_axis.append(i)
        y_axis.append(Fx)

        # apply iteration
        x_old = x
        x = x_old + alpha * d
        x = np.clip(x, -1, 1)

        # Convergence criterion
        if np.linalg.norm(x - x_old) / np.linalg.norm(x_old) < 0.001:
            break

    return x_axis, y_axis
```

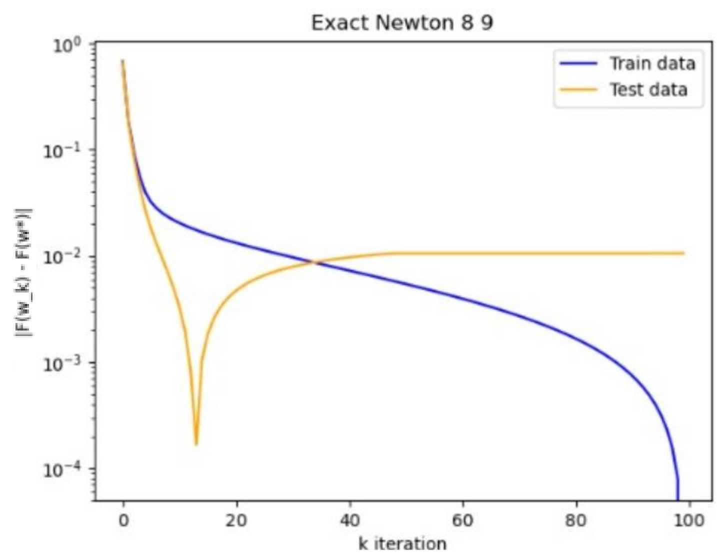
```python
def newton(A, x, labels):
    c1 = np.array(labels)
    x_axis = []
    y_axis = []

    for i in range(100):
        Fx, grad_x, hess_x = func(A, x, c1, True)
        shape = np.shape(hess_x)
        hess_x = hess_x + 0.01 * np.eye(shape[0], shape[1])
        d = -1 * np.linalg.inv(hess_x) @ grad_x
        alpha = linesearch(x, Fx, grad_x, d, 1, 0.5, 1e-4, c1, A)

        x_axis.append(i)
        y_axis.append(Fx)

        # apply iteration
        x_old = x
        x = x_old + alpha * d
        x = np.clip(x, -1, 1)

        # Convergence criterion
        if np.linalg.norm(x - x_old)/np.linalg.norm(x_old) < 0.001:
            break

    return x_axis, y_axis
```

```python
def linesearch(x, Fx, grad_x, d, alpha, beta, c, c1, A):
    for j in range(10):
        Fx_ad, grad_ad = func(A, x + alpha * d, c1)
        if Fx_ad <= Fx + c * alpha * np.dot(grad_x, d):
            break
        else:
            alpha = beta * alpha
    return alpha
```









Without overfitting – change in the change of Hessian to SPD by adding a scaled identity.