Homework 1

In [1]:
```
# Python 3.8.10
%matplotlib inline
```

In [2]:
```
import numpy as np
import os                                    # Only to save the pics in
import matplotlib.pyplot as plt
```

Problem 1

1.1 Inner Product of Vectors

$$< \mathbf{y}, \mathbf{z} >= \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 5 \\ 6 \\ 0 \end{bmatrix} = 6$$

1.2 Inner Product of Matrix and vector

$$\mathbf{X} \cdot \mathbf{y} = \begin{bmatrix} 1 & 2 & 4 \\ 4 & 1 & 2 \\ 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}$$

1.3 Inverse of the Matrix

To see if the matrix $\mathbf{X}$ is invertiable, we can compute its determinant

$$\det(\mathbf{X}) = (2 + 0 + 16) - (16 + 0 + 2) = 0,$$

for the determinant of $\mathbf{X}$ is 0, the matrix has no inverse.

1.4 Solution of a Linear Functions System

The given function can be rewritten as follows:

$$\mathbf{x} = \mathbf{X}^{-1}\mathbf{y} \quad \text{and} \quad \mathbf{x} = \mathbf{X}^{-1}\mathbf{z}$$

as already seen, matrix $\mathbf{X}$ has no inverse, the functions has no solution.

1.5 Eigenvectors and Eigenvalues

Let $\vec{x}$ and λ be the eigenvector and eigenvalue for $\mathbf{Y}$, we have:

$$\begin{aligned} \mathbf{Y} \cdot \vec{x} &= \lambda \cdot \vec{x} \\ \Rightarrow (\mathbf{Y} - \lambda\mathbf{I})\vec{x} &= 0 \\ \Rightarrow \det(\mathbf{Y} - \lambda\mathbf{I}) &= 0 \end{aligned}$$

which leads to:

$$(1-\lambda)^2 - 4 = 0 \quad \Rightarrow \quad \lambda_1 = 3;\ \lambda_2 = -1$$

$$\vec{x}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \vec{x}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

1.6 Partial Derivative and Gradient

The partial derivative of f over x_1 can be given as

$$\begin{aligned} \frac{\partial f(\mathbf{x})}{\partial(x_1)} &= 2(y - \mathbf{z}^\top \mathbf{x}) \cdot \frac{\partial y - \mathbf{z}^\top \mathbf{x}}{\partial x_1} \\ &= 2z_1(\mathbf{z}^\top \mathbf{x} - y) \end{aligned}$$

The gradient of f over $\mathbf{x}$ is

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = 2\mathbf{z}(\mathbf{z}^\top \mathbf{x} - y)$$

1.7 Minimizer

To find the minimizer of f, we have to find the $\mathbf{x}$ such that $\nabla_{\mathbf{x}} f(\mathbf{x}) = 0$

Problem 2

2.1 Mean and Variance

The sample mean is

$$\frac{5}{10} = 0.5$$

and sample variance is

$$\begin{aligned} \text{var} &= \frac{1}{10} \sum_{i=1}^{n} (x_i - 0.5)^2 \\ &= \frac{1}{10} \sum_{i=1}^{10} 0.25^2 \\ &= 0.25 \end{aligned}$$

2.2 Probability of Observation if $p = 0.5$

The probability of observing the given data under the specified circumstance can be calculated as

$$C\binom{10}{5} * 2^{-10} \approx 0.246$$

2.3 p that Maximize the Observation

Let p be the probability to be found, we have:

$$P = p^5 \times (1-p)^5$$

and thus

$$\frac{\partial P}{\partial p} = 5p^4 \times (1-p)^5 \times 5(1-p)^4$$

to solve the problm, let $\frac{\partial P}{\partial p} = 0$, we have

$$\begin{aligned} 0 &= 5p^4 \times (1-p)^5 \times 5(1-p)^4 \\ (1-p)^5 &= p(1-p)^4 \\ (1-p) &= p \\ p &= 0.5 \end{aligned}$$

2.4 Distribution of a Combined Variable

$X + aY$ is still a Gaussian random variable

2.5 Inpendency Verification of a Combined Variable

The new variable $Z = XY$ is independent on Y.

Because Y is a random varible with $Y \in \{1, -1\}$, if $Y = 1$ it is intuitive that $Z = X$. If $Y = -1$, even though we have $Z = -X$, but since $X \sim \mathcal{N}(0, \sigma^2)$, X is symmetric around 0, we can still say that $Z = X$.

Problem 3

3.1

(a)

Both, the answer depends on the choice of C, to the given functions we can apply the change of base fomular for logarithms, which gives us

$$\log_2(x) = \frac{\log_e(x)}{\log_e(2)},$$

indicating if $C = \frac{1}{\log_e(2)}$, then $f(x) = O(g(x))$, and if $C = \log_e(2)$, $g(x) = O(f(x))$

(b)

Assuming $f(x) = O(g(x))$, which gives us

$$\begin{aligned} |e^x| &\le C2^x \\ \log_e(e^x) &\le \log_e(C2^x) \\ x &\le \log_e(C2^x) \\ &\le \log_e C + log_e 2^x \\ &\le \log_e C + xlog_e 2 \end{aligned}$$

for $log_e 2 < 1$, there exists no valid x_0 and C.

Likely we can also asumme $g(x) = O(f(x))$, and the following should hold

$$\begin{aligned} |2^x| &\leq Ce^x \\ \log_2(2^x) &\leq \log_2(Ce^x) \\ x &\leq \log_2(Ce^x) \\ &\leq \log_2 C + log_2 e^x \\ &\leq \log_2 C + x log_2 e \end{aligned}$$

The inequality can be true if we simply set $C = 1$ because $\log_2 e > 1$

Thus for this task, $g(x) = O(f(x))$.

(c)

Assuming $f(x) = O(g(x))$, which gives us

$$\begin{aligned} |x| &\leq C(10x + \log_2 x) \\ &\leq (10Cx + C\log_2 x) \end{aligned}$$

valid solution can be found if we simply set $C = 1$, so $f(x) = O(g(x))$.

3.2

The calculation goes:

$$\mathbf{X}\mathbf{y} \mapsto \mathbb{R}^{m \times n} \times \mathbb{R}^n$$

in every array of $\mathbf{X}$ there are n multiplications and $n - 1$ additions, and we have m such procedures, so we have the complexity of $O(m(2n - 1))$

Problem 4

4.1 Scatter Plot

create the required x array

In [3]:
```
x = np.arange(0, 1, 0.001)
print("Size of the x:", x.size)
```

```
Size of the x: 1000
```

genetate ϵ_i following gaussian distribution

In [4]:
```
epsilon = np.random.normal(0, 0.25, 1000)
print("Size of the epsilon:", epsilon.size)
```

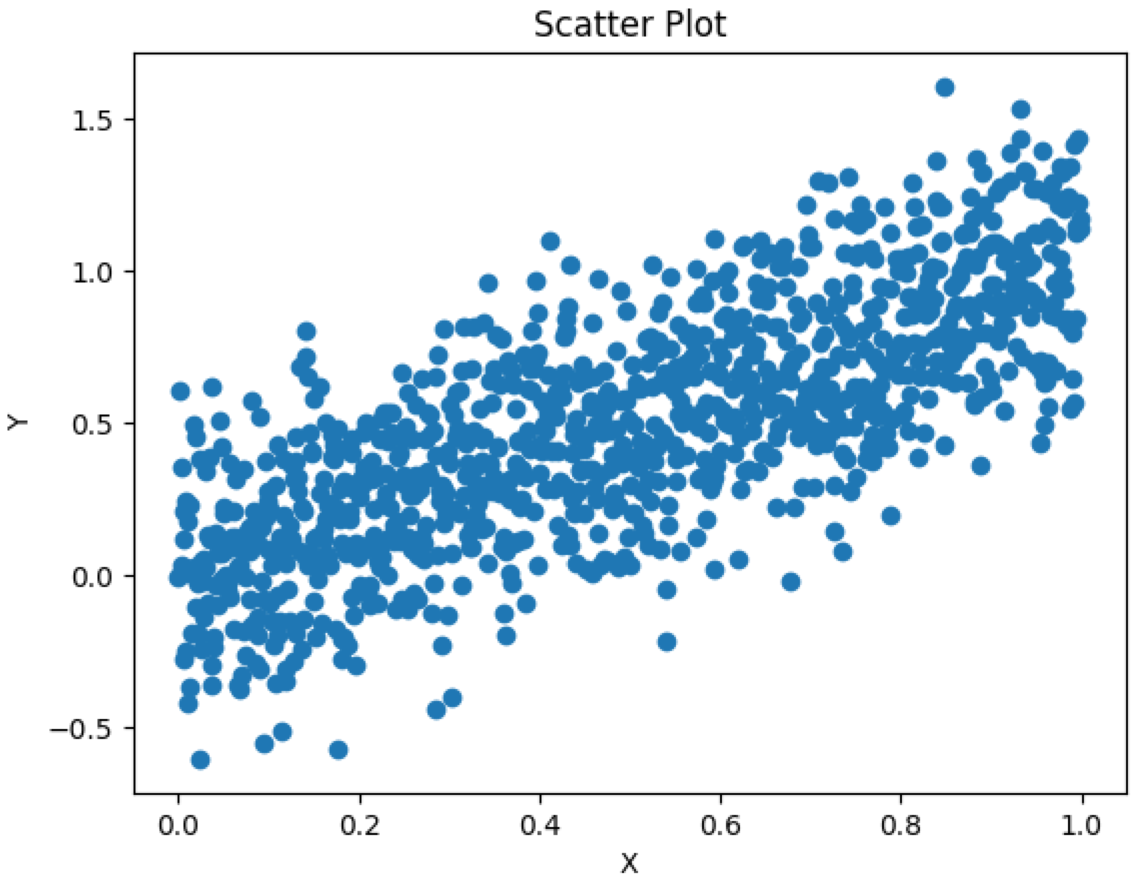
```
Size of the epsilon: 1000
```

compute y

In [5]:
```
y = x + epsilon
```

make the scatter plot

In [6]:
```
plt.scatter(x, y)
plt.xlabel('X')
plt.ylabel('Y')
plt.title('Scatter Plot')
plt.savefig("/home/yueyangzhang/TUM/MLO/Problem4_1.png")
```



4.2 Compute a and Plot

The minimum of $f(a) = \sum_{i=1}^{n}(x_i a - y_i)^2$ can be represented as $\text{argmin}_a \sum_{i=1}^{n}(x_i a - y_i)^2$, regarding to proposition 1 in lecture_note_1_2, a can be calculated as $(\underline{x}^\top \underline{x})^{-1}\underline{x}^\top \underline{y}$

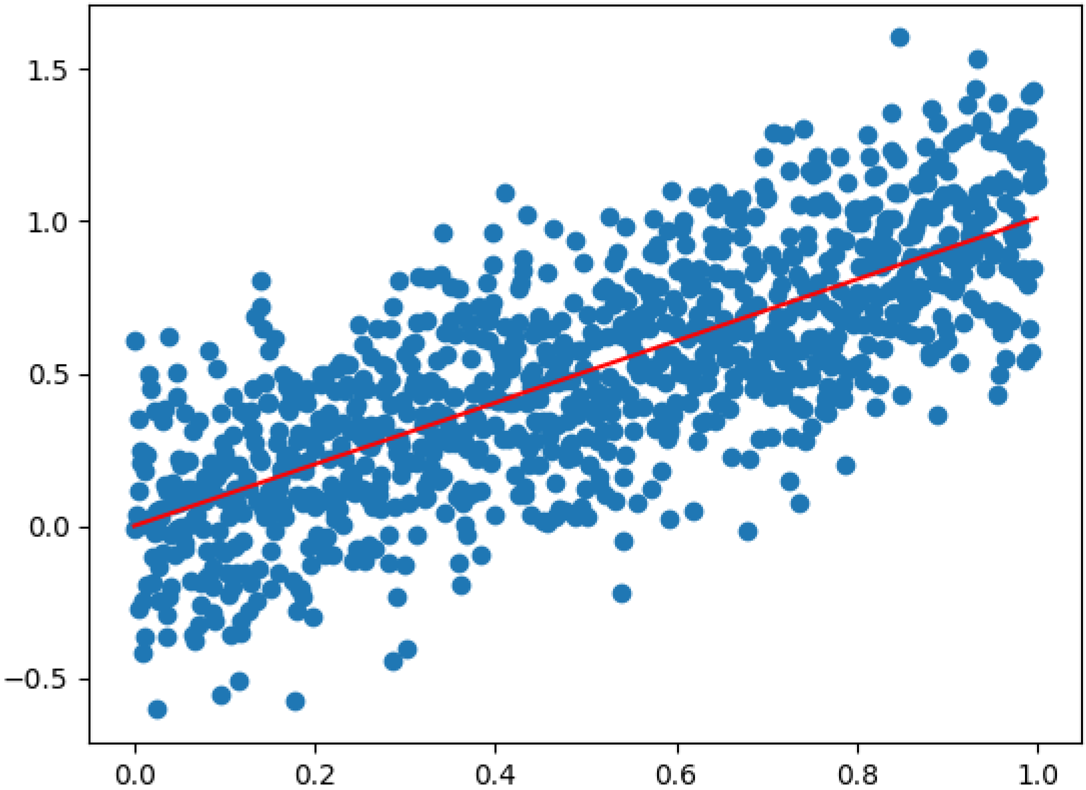
In [7]:
```
a = np.dot(pow(np.dot(x, x), -1), np.dot(x, y))
print("minimizer of f w.r.t. a is a =", a)
```

```
minimizer of f w.r.t. a is a = 1.0114032989249062
```

add the estimated linear model to the data

In [8]:
```
g = np.dot(a, x)
plt.scatter(x, y)
plt.plot(x, g, color = "red")
# plt.savefig("/home/yueyangzhang/TUM/MLO/Problem4_2.png")
```

Out[8]: `[<matplotlib.lines.Line2D at 0x7fa902629400>]`



4.3 Close Form Solution

The gievn function can be rewritten as:

$$p_d(\mathbf{x}) = \mathbf{X} \cdot \mathbf{a}$$

where

$$\mathbf{X} = \begin{bmatrix} x_1^0 & \cdots & x_1^d \\ \vdots & \ddots & \vdots \\ x_n^0 & \cdots & x_n^d \end{bmatrix} \in \mathbb{R}^{n\times(d+1)}$$

and $a = [a_0 \ \cdots \ a_d]^\top \in \mathbb{R}^{d+1}$. Consequrntly, $f(\mathbf{a})$ can be written as:

$$f(\mathbf{a}) = \left\|\mathbf{X}\mathbf{a}^\top - y\right\|_2^2$$

and thus the close form solution can be refered to the lecture note and given as

$$\mathbf{a} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top y$$

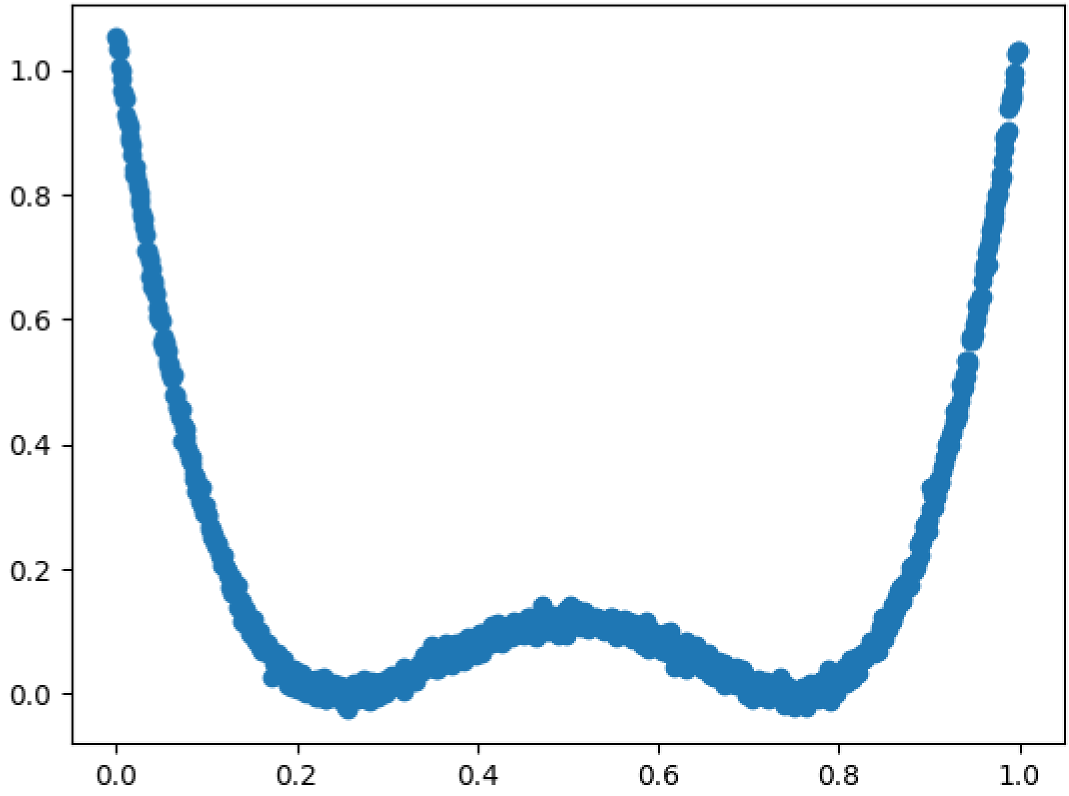
4.4 Upgraded y and Plot

as requested change the ground truth of y, denoted as y_1

In [9]:
```
xi = np.random.normal(0, 0.01, 1000)

y_1 =  30 * ((x - 0.25) ** 2 * (x - 0.75) ** 2) + xi
plt.scatter(x, y_1)
```

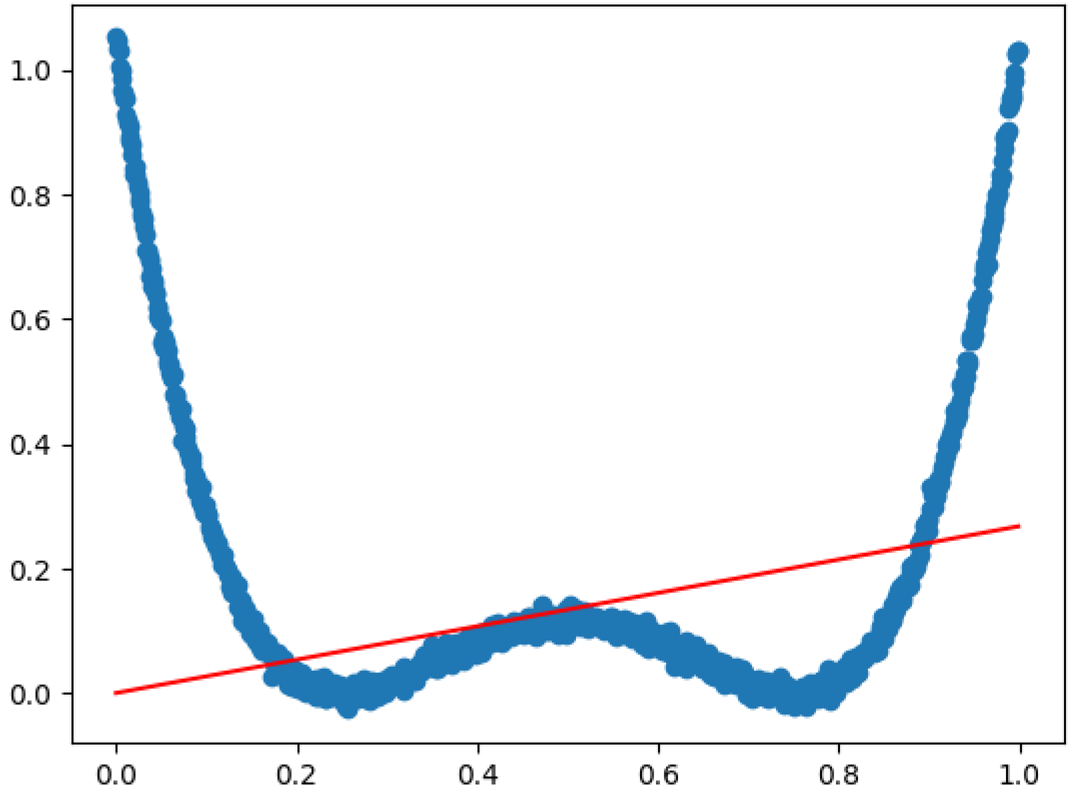
Out[9]: <matplotlib.collections.PathCollection at 0x7fa94414e2e0>



For the new curve, applies the closed form of the least square solution

In [10]:
```
a_1 = np.dot(pow(np.dot(x, x), -1), np.dot(x, y_1))
g_1 = a_1 * x
plt.scatter(x, y_1)
plt.plot(x, g_1, color = "red")
# plt.savefig("/home/yueyangzhang/TUM/MLO/Problem4_4.png")
```

Out[10]: [<matplotlib.lines.Line2D at 0x7fa9024b87f0>]



Problem 5

5.1 Perpandicularity Proof

Set $b = 0$, we can get $< \mathbf{w}, \mathbf{x} >= 0$, in this case $\mathbf{w} \perp \mathbf{x}$. We can further translate $\mathbf{x}$ by an arbitrary distance $\mathbf{a}$, because this is just a pure translation, we get $< \mathbf{w}, (\mathbf{x} + \mathbf{a}) >= 0$, alternatively

$$\mathbf{w}^\top (\mathbf{x} + \mathbf{a}) = 0$$
$$\Rightarrow \mathbf{w}^\top \mathbf{x} + \mathbf{w}^\top \mathbf{a} = 0,$$

for that $\mathbf{a}$ is arbitrary, we can let $b = \mathbf{w}^\top \mathbf{a}$, the result is still a translation of $\mathbf{x}$, so the relationship holds $\mathbf{w} \perp \mathbf{x}$

5.2 Distance

Follow the steps in 5.1 and w.l.o.g. let $\mathbf{a}$ perpandicular to the hyperplane, which is the shortest vector between $\mathbf{w}$ and $\mathbf{x}$, the distance can be calculated as $\text{distance} = ||\mathbf{a}||$, i.e.,

$$||\mathbf{a}|| = \frac{||\mathbf{b}||}{||\mathbf{w}||}$$

Problem 6

6.1 Variance

$$\begin{aligned}\mathrm{Var}(Y) &= a^2 \times \mathrm{Var}(X) \\ &= a^2\sigma^2\end{aligned}$$

6.2 (a) CLT

If n is large enough, according to the Central Limit Theorem (CLT), Y is converging to a standard normal distribution, i.e. $Y \sim \mathcal{N}(0, \sigma^2)$.

6.2 (b) Interval

Follow the result of 6.2 (a) we have $Y \sim \mathcal{N}(0, 0.25)$ and thus $Y \sim \mathcal{N}(0, 1)$. For such a distribution, 95% lies in the inverval $(-1.96, 1.96)$. For the solution, the interval is given as

$$\frac{1}{n}(\frac{n}{2} \pm 1.96 \cdot \frac{\sqrt{n}}{2}) \quad \Rightarrow \quad (0.490.51)$$

6.3 Test and Sickness

Define the following events

1. A: person is sick
2. B: test result is positive
3. C: person is health
4. D: test result is negative

and according to the given information we have $P(A) = 0.5\%$, $P(B|A) = 95\%$ and $P(B|C) = 10\%$

(a)

To be found is $P(B)$

$$\begin{aligned}PB &= P(B|A)P(A) + P(B|C)P(C) \\ &= 0.005 \times 0.95 + (1 - 0.005) \times 0.1 \\ &\approx 10\%\end{aligned}$$

(b)

To be found is $P(A|B)$, we can utilize Bayes' rule

$$\begin{aligned}PA|B &= \frac{P(B|A)P(A)}{P(B)} \\ &= \frac{0.95 \times 0.005}{0.1} \\ &\approx 4\%\end{aligned}$$

(c)

To be found is $P(C|D)$

$$
\begin{aligned}
P(C|D) &= 1 - P(A|D) \\
&= 1 - \frac{P(D|A)P(A)}{P(D)} &= 1 - \frac{(1 - P(B|A))P(A)}{1 - P(B)} \approx 99\%
\end{aligned}
$$

(d)

To be found can be formulated as

$$
\begin{aligned}
P(B|C)P(C) + P(D|A)P(A) &= 0.1(1 - 0.005) + (1 - 0.95)0.005 \\
&\approx 10\%
\end{aligned}
$$

Problem 7

7.1 Global Minimum of Convex Function

Let $\mathbf{x}^*$ be the local minimum of f, assume it is not the global minimum, which meams there exist a $\mathbf{x}_0$ such that $f(\mathbf{x}_0) < f(\mathbf{x}^*)$.

By definition, for a convex function, for any $\theta \in (0, 1)$ we have

$$
\begin{aligned}
f(\theta\mathbf{x}^* + (1-\theta)\mathbf{x}_0) &\leq \theta f(\mathbf{x}^*) + (1-\theta) f(\mathbf{x}_0) \\
&< \theta f(\mathbf{x}^*) + (1-\theta) f(\mathbf{x}^*) \\
\Rightarrow f(\theta\mathbf{x}^* + (1-\theta)\mathbf{x}_0) &< f(\mathbf{x}^*)
\end{aligned}
$$

Now, if $\theta \to 1$, $\theta\mathbf{x}^* + (1-\theta)\mathbf{x}_0 \to \mathbf{x}^*$, in such case, $\mathbf{x}^*$ cannot be the local minimum neither.

Therefore local minimun $\mathbf{x}^*$ of a convex function must also be the global minimum.