LINEAR REGRESSION: MLE AND MAP

Matee Vadrukchid

LIST OUTLINE

- Derivative

  https://ocw.mit.edu/courses/res-18-001-calculus-fall-2023/pages/textbook/

- Linear Regression

DERIVATIVE

Question 1:

$$y = 2x^4 - 7x^3$$

What is X that leads to the minimum Y?

1. $\frac{dc}{dx} = 0$ $\qquad c' = 0$
2. $\frac{d}{dx}(x^n) = nx^{n-1}$ $\qquad (x^n)' = nx^{n-1}$
3. $\frac{d}{dx}(f+g) = \frac{df}{dx} + \frac{dg}{dx}$ $\qquad (f+g)' = f' + g'$
4. $\frac{d}{dx}(cf+kg) = c\frac{df}{dx} + k\frac{dg}{dx}$ $\qquad (cf+kg)' = cf' + kg'$
5. $\frac{d}{dx}(fg) = \frac{df}{dx}g + \frac{dg}{dx}f$ $\qquad (fg)' = f'g + g'f$
6. $\frac{d}{dx}(\frac{f}{g}) = \frac{g(x)\frac{d}{dx}f(x) - f(x)\frac{d}{dx}g(x)}{g(x)^2}$ $\qquad (\frac{f}{g})' = \frac{gf' - fg'}{g^2}$
7. $\frac{d}{dx}(u^n) = nu^{n-1}\frac{du}{dx}$
8. $\frac{d}{dx}(f \circ g) = \frac{d}{dg(x)}f(g(x))\frac{d}{dx}g(x)$ $\qquad (f \circ g)'(x) = f'(g(x))g'(x)$
9. $\frac{d}{dx}\ln|x| = \frac{1}{x}$
10. $\frac{d}{dx}e^x = e^x$
11. $\frac{d}{dx}a^x = a^x \ln a$
12. $\frac{d}{dx}\log_a|x| = \frac{1}{x \ln a}$
13. $\frac{d}{dx}\sin x = \cos x$
14. $\frac{d}{dx}\cos x = -\sin x$
15. $\frac{d}{dx}\tan x = \sec^2 x$
16. $\frac{d}{dx}\cot x = -\mathrm{cosec}^2 x$
17. $\frac{d}{dx}\sec x = \sec x \tan x$
18. $\frac{d}{dx}\mathrm{cosec}\, x = -\mathrm{cosec}\, x \cot x$
19. $\frac{d}{dx}\arcsin x = \frac{1}{\sqrt{1-x^2}}$
20. $\frac{d}{dx}\arccos x = -\frac{1}{\sqrt{1-x^2}}$
21. $\frac{d}{dx}\arctan x = \frac{1}{1+x^2}$
22. $\frac{d}{dx}\mathrm{arccot}\, x = -\frac{1}{1+x^2}$
23. $\frac{d}{dx}\mathrm{arcsec}\, x = \frac{1}{|x|\sqrt{x^2-1}}$
24. $\frac{d}{dx}\mathrm{arccosec}\, x = -\frac{1}{|x|\sqrt{x^2-1}}$

DERIVATIVE

Question 1:

$$y' = 8x^3 - 21x^2$$

$$8x^3 - 21x^2 = 0$$

$$x^2(8x - 21) = 0$$

$$x = 0 \quad \text{or} \quad x = \frac{21}{8} \approx 2.625$$

DERIVATIVE

Question 1:

$$y'' = 24x^2 - 42x$$

$y''(0) = 24 \times 0^2 - 42 \times 0 = 0$ NEED PROOF FURTHER

$y''\left(\frac{21}{8}\right) = 24\left(\frac{21}{8}\right)^2 - 42\left(\frac{21}{8}\right) = 24 \times \frac{441}{64} - 42 \times \frac{21}{8} = \frac{10584}{64} - \frac{882}{8} = \frac{10584-7056}{64} = \frac{3528}{64} = 55.125$ > 0

At $x = 0$, $y = 0$

At $x = \frac{21}{8}$, $y \approx -31.65$

DERIVATIVE

Question 2:

$$y = -3x^4 + 4x^3$$

What is X that leads to the maximum Y?

1. $\frac{dc}{dx} = 0$ $\qquad c' = 0$
2. $\frac{d}{dx}(x^n) = nx^{n-1}$ $\qquad (x^n)' = nx^{n-1}$
3. $\frac{d}{dx}(f + g) = \frac{df}{dx} + \frac{dg}{dx}$ $\qquad (f + g)' = f' + g'$
4. $\frac{d}{dx}(cf + kg) = c\frac{df}{dx} + k\frac{dg}{dx}$ $\qquad (cf + kg)' = cf' + kg'$
5. $\frac{d}{dx}(fg) = \frac{df}{dx}g + \frac{dg}{dx}f$ $\qquad (fg)' = f'g + g'f$
6. $\frac{d}{dx}(\frac{f}{g}) = \frac{g(x)\frac{d}{dx}f(x) - f(x)\frac{d}{dx}g(x)}{g(x)^2}$ $\qquad (\frac{f}{g})' = \frac{gf' - fg'}{g^2}$
7. $\frac{d}{dx}(u^n) = nu^{n-1}\frac{du}{dx}$
8. $\frac{d}{dx}(f \circ g) = \frac{d}{dg(x)}f(g(x))\frac{d}{dx}g(x)$ $\qquad (f \circ g)'(x) = f'(g(x))g'(x)$
9. $\frac{d}{dx}\ln|x| = \frac{1}{x}$
10. $\frac{d}{dx}e^x = e^x$
11. $\frac{d}{dx}a^x = a^x \ln a$
12. $\frac{d}{dx}\log_a|x| = \frac{1}{x\ln a}$
13. $\frac{d}{dx}\sin x = \cos x$
14. $\frac{d}{dx}\cos x = -\sin x$
15. $\frac{d}{dx}\tan x = \sec^2 x$
16. $\frac{d}{dx}\cot x = -\operatorname{cosec}^2 x$
17. $\frac{d}{dx}\sec x = \sec x \tan x$
18. $\frac{d}{dx}\operatorname{cosec} x = -\operatorname{cosec} x \cot x$
19. $\frac{d}{dx}\arcsin x = \frac{1}{\sqrt{1-x^2}}$
20. $\frac{d}{dx}\arccos x = -\frac{1}{\sqrt{1-x^2}}$
21. $\frac{d}{dx}\arctan x = \frac{1}{1+x^2}$
22. $\frac{d}{dx}\operatorname{arccot} x = -\frac{1}{1+x^2}$
23. $\frac{d}{dx}\operatorname{arcsec} x = \frac{1}{|x|\sqrt{x^2-1}}$
24. $\frac{d}{dx}\operatorname{arccosec} x = -\frac{1}{|x|\sqrt{x^2-1}}$

DERIVATIVE

Question 2:

$$y' = \frac{d}{dx}(-3x^4 + 4x^3) = -12x^3 + 12x^2$$

$$-12x^3 + 12x^2 = 0$$

$$12x^2(-x + 1) = 0$$

$$x = 0 \quad \text{or} \quad x = 1$$

DERIVATIVE

Question 2:

$$y'' = \frac{d}{dx}(-12x^3 + 12x^2) = -36x^2 + 24x$$

$$y''(0) = -36(0)^2 + 24(0) = 0$$

$$y''(1) = -36(1)^2 + 24(1) = -36 + 24 = -12 < 0$$

$$y = -3(1)^4 + 4(1)^3 = -3 + 4 = 1$$

DERIVATIVE

Question 3:

The derivative of e^{3x} ❓

10. $\frac{d}{dx}e^x = e^x$

CHAIN RULE...

DERIVATIVE

Question 4:

The derivative of $e^{(\ln 2)x}$ ❓

CHAIN RULE...

9. $\frac{d}{dx}\ln|x| = \frac{1}{x}$

10. $\frac{d}{dx}e^x = e^x$

11. $\frac{d}{dx}a^x = a^x \ln a$

12. $\frac{d}{dx}\log_a|x| = \frac{1}{x\ln a}$

Properties of natural logs

1. $e^{\ln x} = x$

Its derivative is $\ln 2$ times 2^x

DERIVATIVE

Question 5:

The derivative of $e^{-x^2/2}$ ❓

CHAIN RULE...

1. $\frac{dc}{dx} = 0$ $\qquad c' = 0$
2. $\frac{d}{dx}(x^n) = nx^{n-1}$ $\qquad (x^n)' = nx^{n-1}$
3. $\frac{d}{dx}(f+g) = \frac{df}{dx} + \frac{dg}{dx}$ $\qquad (f+g)' = f' + g'$
4. $\frac{d}{dx}(cf + kg) = c\frac{df}{dx} + k\frac{dg}{dx}$ $\qquad (cf + kg)' = cf' + kg'$
5. $\frac{d}{dx}(fg) = \frac{df}{dx}g + \frac{dg}{dx}f$ $\qquad (fg)' = f'g + g'f$
6. $\frac{d}{dx}(\frac{f}{g}) = \frac{g(x)\frac{d}{dx}f(x) - f(x)\frac{d}{dx}g(x)}{g(x)^2}$ $\qquad (\frac{f}{g})' = \frac{gf' - fg'}{g^2}$
7. $\frac{d}{dx}(u^n) = nu^{n-1}\frac{du}{dx}$
8. $\frac{d}{dx}(f \circ g) = \frac{d}{dg(x)}f(g(x))\frac{d}{dx}g(x)$ $\qquad (f \circ g)'(x) = f'(g(x))g'(x)$
9. $\frac{d}{dx}\ln|x| = \frac{1}{x}$
10. $\frac{d}{dx}e^x = e^x$

DERIVATIVE

Question 6:

The derivative of $y = \ln\left(\frac{x+2}{x-1}\right)$ ❓

CHAIN RULE...

1. $\frac{dc}{dx} = 0$ $\quad c' = 0$
2. $\frac{d}{dx}(x^n) = nx^{n-1}$ $\quad (x^n)' = nx^{n-1}$
3. $\frac{d}{dx}(f+g) = \frac{df}{dx} + \frac{dg}{dx}$ $\quad (f+g)' = f' + g'$
4. $\frac{d}{dx}(cf+kg) = c\frac{df}{dx} + k\frac{dg}{dx}$ $\quad (cf+kg)' = cf' + kg'$
5. $\frac{d}{dx}(fg) = \frac{df}{dx}g + \frac{dg}{dx}f$ $\quad (fg)' = f'g + g'f$
6. $\frac{d}{dx}\left(\frac{f}{g}\right) = \frac{g(x)\frac{d}{dx}f(x) - f(x)\frac{d}{dx}g(x)}{g(x)^2}$ $\quad \left(\frac{f}{g}\right)' = \frac{gf' - fg'}{g^2}$
7. $\frac{d}{dx}(u^n) = nu^{n-1}\frac{du}{dx}$
8. $\frac{d}{dx}(f \circ g) = \frac{d}{dg(x)}f(g(x))\frac{d}{dx}g(x)$ $\quad (f \circ g)'(x) = f'(g(x))g'(x)$
9. $\frac{d}{dx}\ln|x| = \frac{1}{x}$
10. $\frac{d}{dx}e^x = e^x$
11. $\frac{d}{dx}a^x = a^x \ln a$
12. $\frac{d}{dx}\log_a|x| = \frac{1}{x\ln a}$

Rules of Logarithms

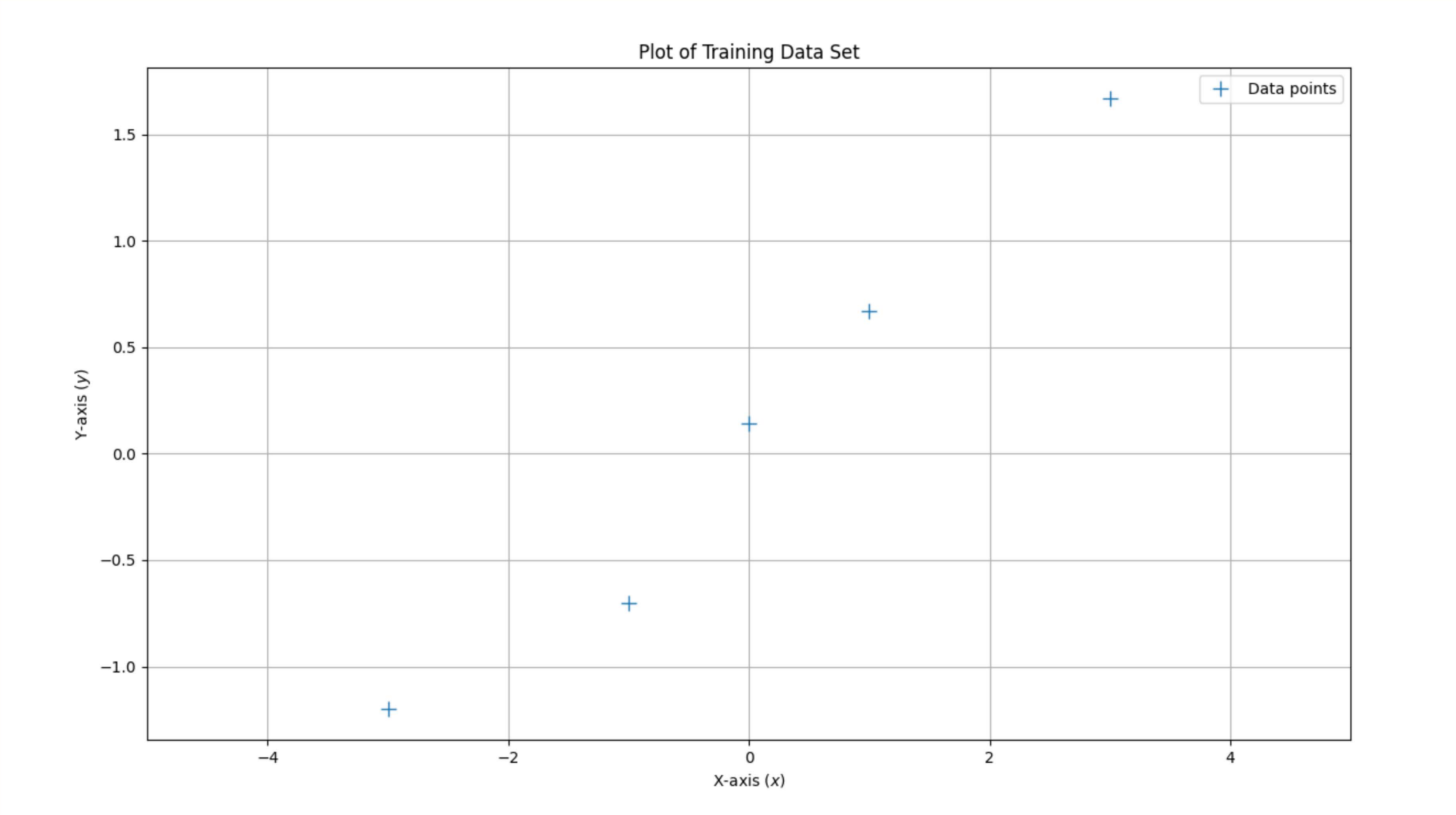
Rule 1: $\log_b(M \cdot N) = \log_b M + \log_b N$

Rule 2: $\log_b\left(\frac{M}{N}\right) = \log_b M - \log_b N$

Rule 3: $\log_b\left(M^k\right) = k \cdot \log_b M$

LINEAR REGRESSION

LINEAR REGRESSION



LINEAR REGRESSION

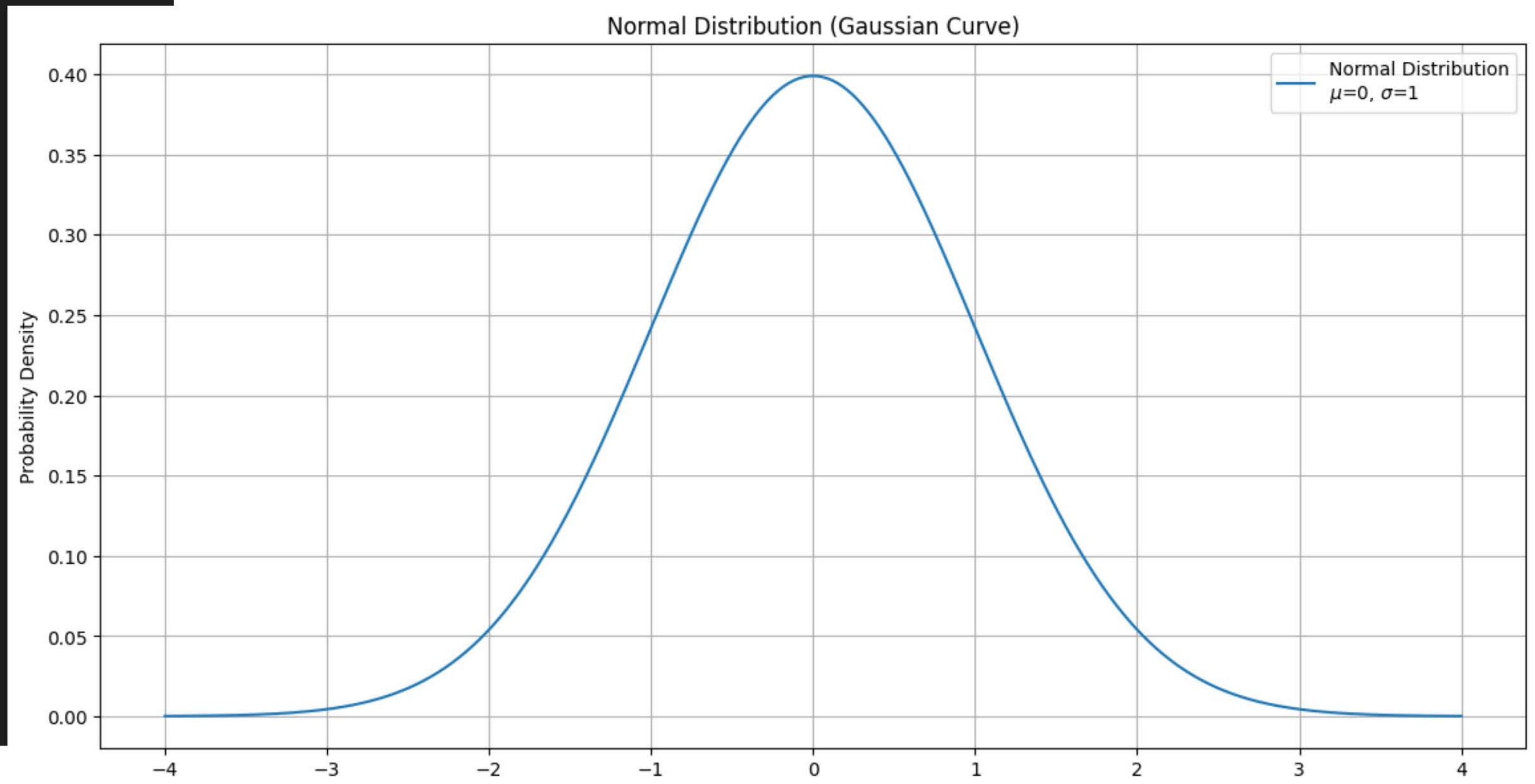
Maximum Likelihood Estimation

```python
import numpy as np
import matplotlib.pyplot as plt

# Define the normal distribution function
def normal_distribution(x, mu, sigma):
    """Compute the normal distribution (Gaussian) for given mean (mu) and standard deviation (sigma)."""
    return (1 / (np.sqrt(2 * np.pi) * sigma)) * np.exp(-0.5 * ((x - mu) / sigma) ** 2)

# Parameters for the normal distribution
mu = 0     # Mean
sigma = 1 # Standard deviation

# Generate x values
x = np.linspace(mu - 4*sigma, mu + 4*sigma, 1000)

# Compute y values using the normal distribution function
y = normal_distribution(x, mu, sigma)

# Plot the normal distribution
plt.figure(figsize=(8, 5))
plt.plot(x, y, label=f"Normal Distribution\n$\mu$={mu}, $\sigma$={sigma}")
plt.title("Normal Distribution (Gaussian Curve)")
plt.xlabel("x")
plt.ylabel("Probability Density")
plt.legend()
plt.grid()
plt.show()
```

NORMAL DISTRIBUTION FUNCTION

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}$$

$\mu =$ Mean
$\sigma =$ Standard Deviation
$\pi \approx 3.14159\cdots$
$e \approx 2.71828\cdots$

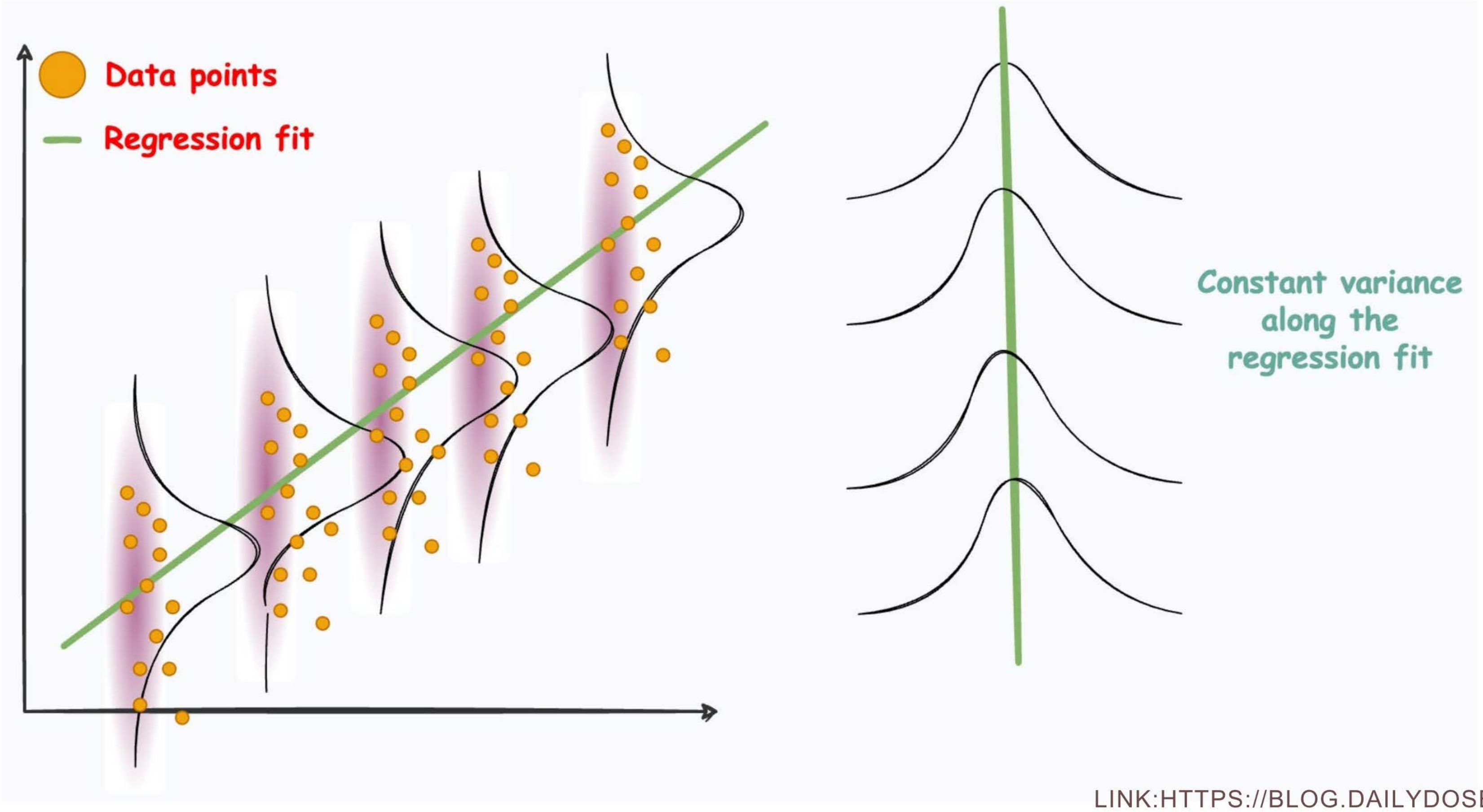


LINEAR REGRESSION

Maximum Likelihood Estimation

$$p(y \mid \boldsymbol{x}) = \mathcal{N}\big(y \mid f(\boldsymbol{x}),\, \sigma^2\big)\,.$$

$$y = f(\boldsymbol{x}) + \epsilon\,, \quad \epsilon \sim \mathcal{N}\big(0,\, \sigma^2\big)\,,$$



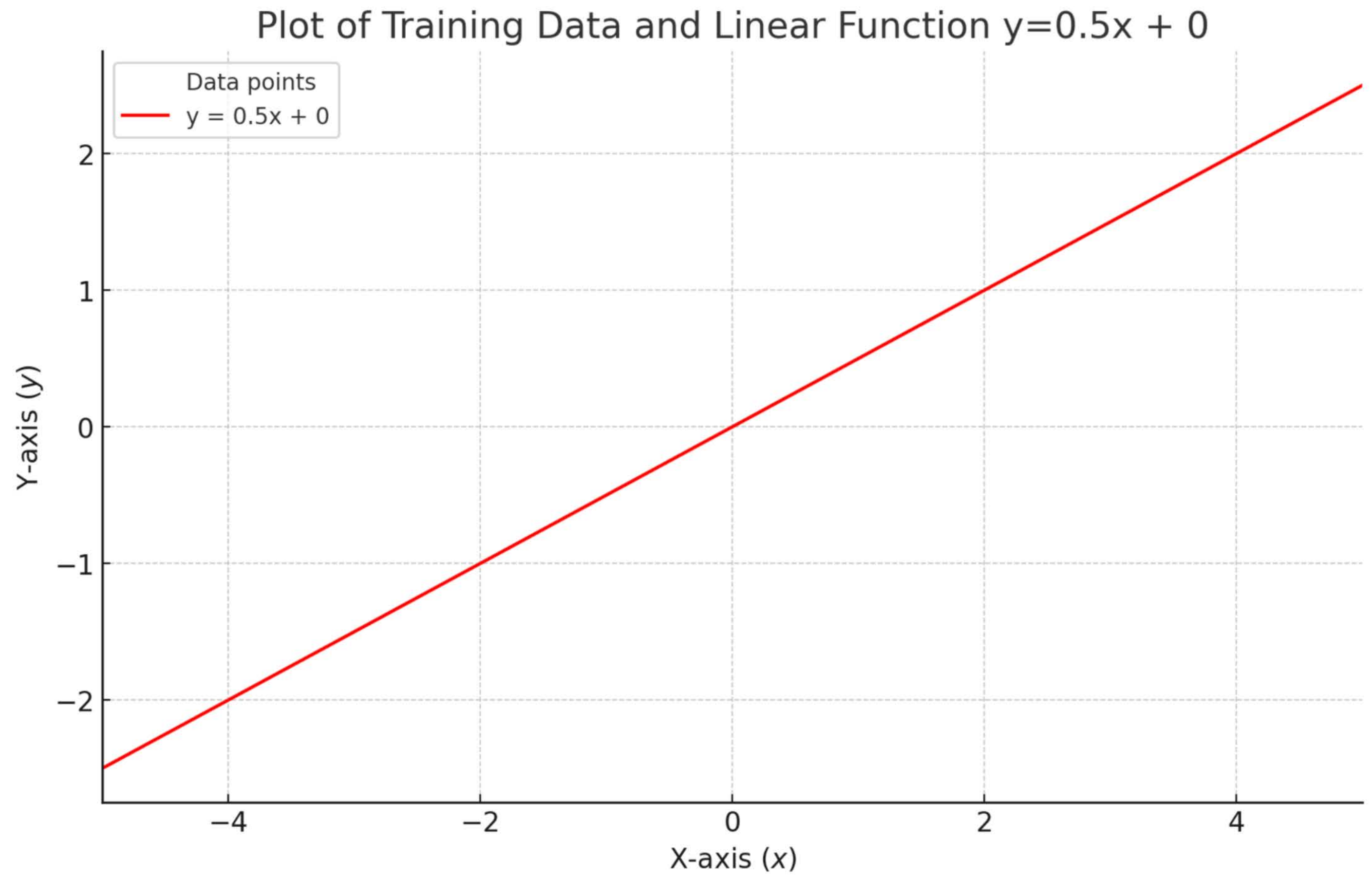
LINK:HTTPS://BLOG.DAILYDOSEOFDS.COM/P/GENERALIZED-LINEAR-MODELS-GLMS-THE

LINEAR REGRESSION

Maximum Likelihood Estimation

$$p(y \mid \boldsymbol{x}, \boldsymbol{\theta}) = \mathcal{N}(y \mid \boldsymbol{x}^\top \boldsymbol{\theta}, \sigma^2) \iff y = \boldsymbol{x}^\top \boldsymbol{\theta} + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma^2)$$

$$y = 0.5x + 0$$

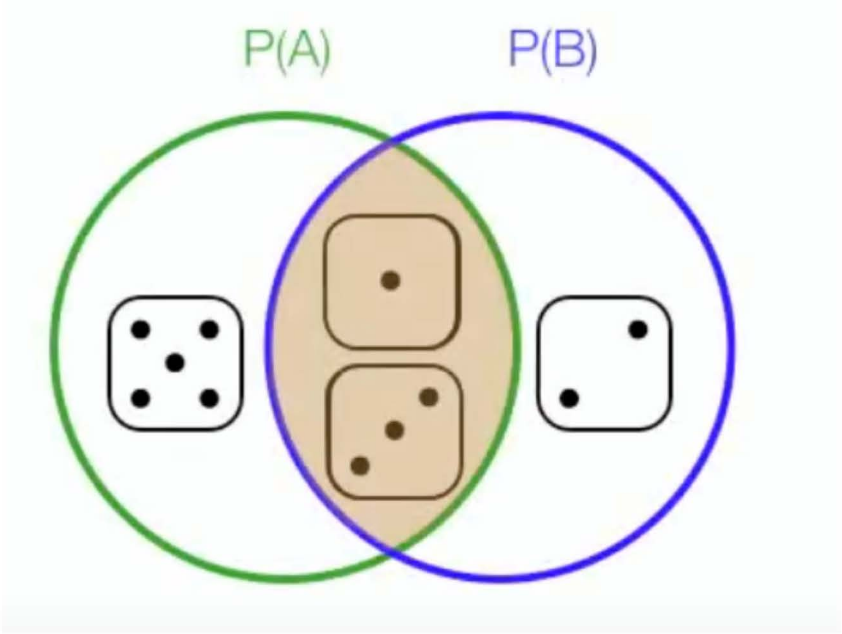


LINEAR REGRESSION

Maximum Likelihood Estimation

Conditional Probability

P(B | A) = What is the Probability of (rolling a dice and it's value is less than 4 | knowing that the value is an odd number)

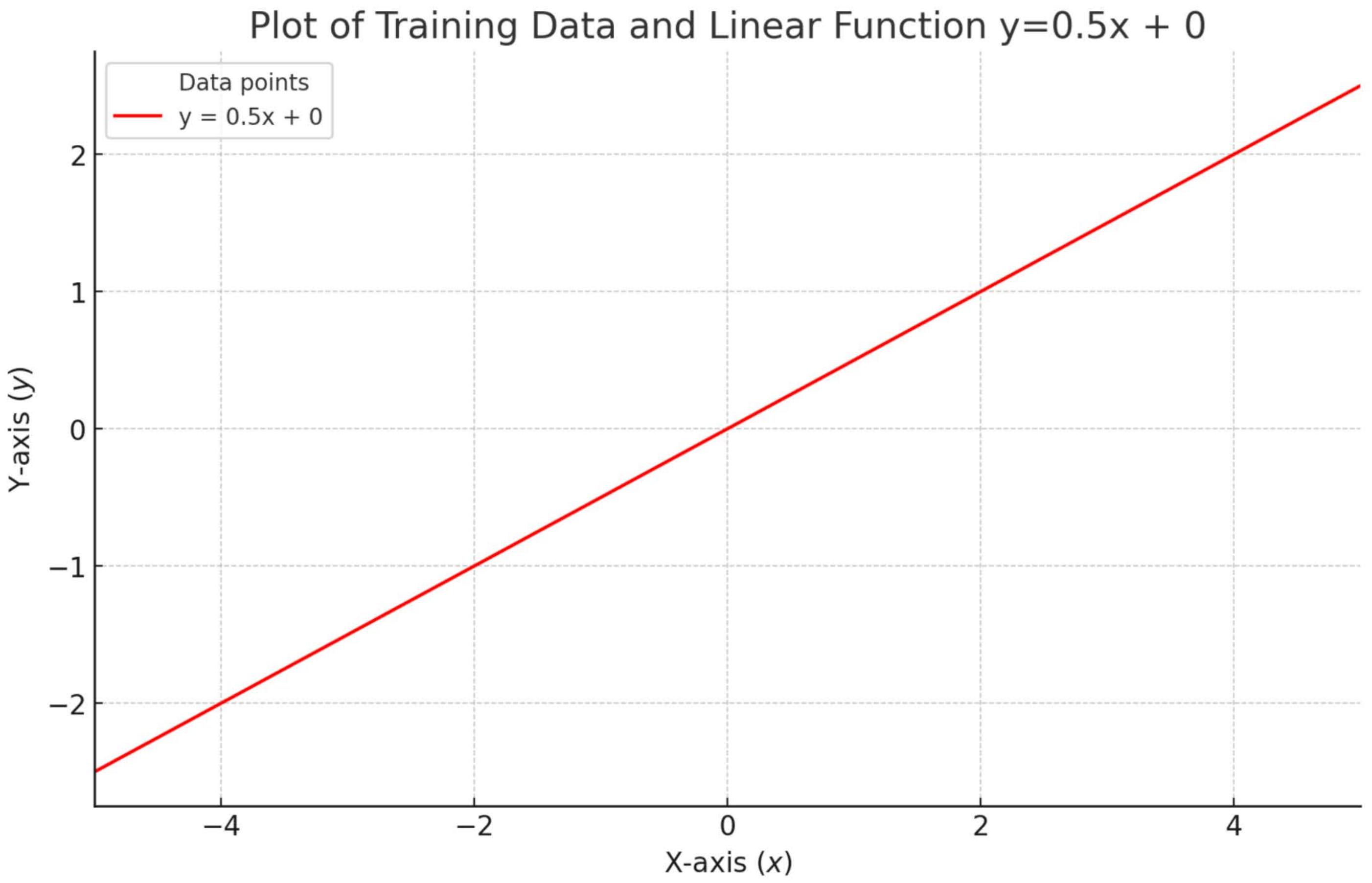


LINEAR REGRESSION

Maximum Likelihood Estimation

$$p(y \mid \boldsymbol{x}, \boldsymbol{\theta}) = \mathcal{N}(y \mid \boldsymbol{x}^\top \boldsymbol{\theta},\, \sigma^2) \iff y = \boldsymbol{x}^\top \boldsymbol{\theta} + \epsilon\,, \quad \epsilon \sim \mathcal{N}(0,\, \sigma^2)$$

$$y = 0.5x + 0$$

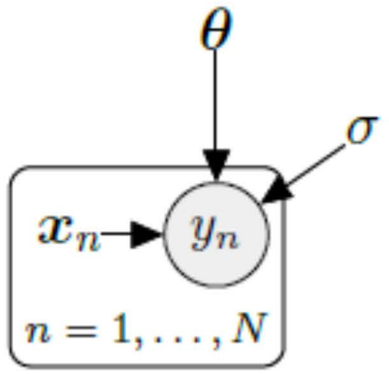


LINEAR REGRESSION

Maximum Likelihood Estimation

$$p(\mathcal{Y} \,|\, \mathcal{X}, \boldsymbol{\theta}) = p(y_1, \ldots, y_N \,|\, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N, \boldsymbol{\theta})$$

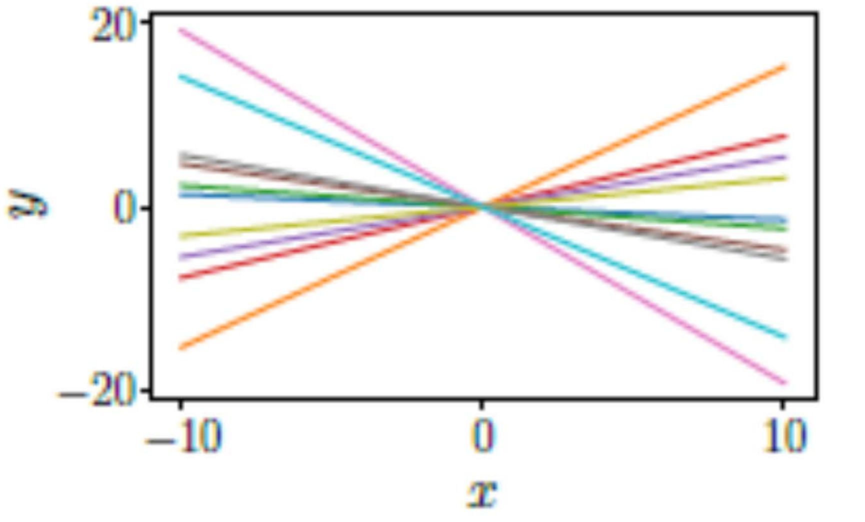
$$= \prod_{n=1}^{N} p(y_n \,|\, \boldsymbol{x}_n, \boldsymbol{\theta}) = \prod_{n=1}^{N} \mathcal{N}\big(y_n \,|\, \boldsymbol{x}_n^\top \boldsymbol{\theta},\, \sigma^2\big)$$



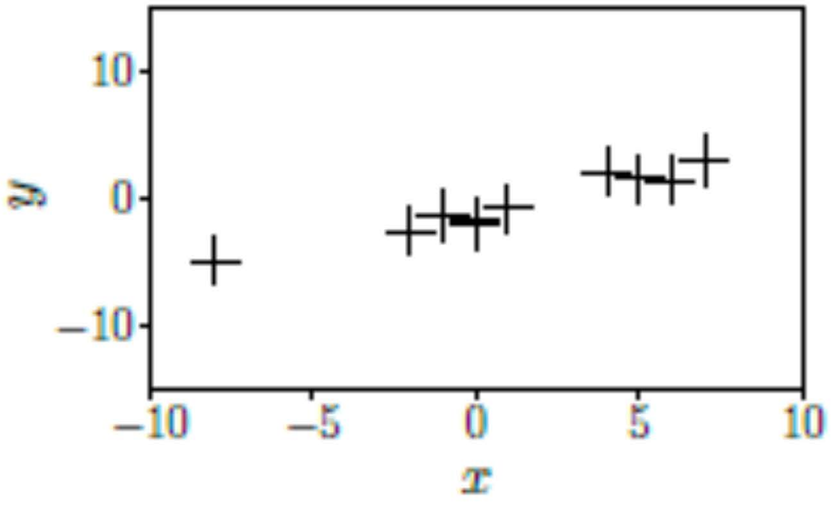
we defined $\mathcal{X} := \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ and $\mathcal{Y} := \{y_1, \ldots, y_N\}$

LINEAR REGRESSION

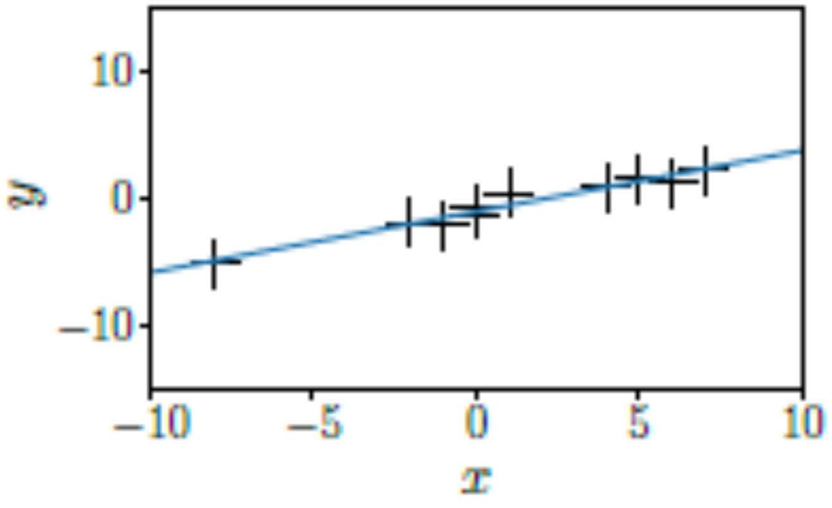
Maximum Likelihood Estimation



(a) Example functions (straight lines) that can be described using the linear model in (9.4).



(b) Training set.



(c) Maximum likelihood estimate.

LINEAR REGRESSION

Maximum Likelihood Estimation

$$\boldsymbol{\theta}_{\mathrm{ML}} \in \arg\max_{\boldsymbol{\theta}} p(\mathcal{Y} \,|\, \mathcal{X}, \boldsymbol{\theta})$$

$$p(\mathcal{Y} \,|\, \mathcal{X}, \boldsymbol{\theta}) = \prod_{n=1}^{N} p(y_n \,|\, \boldsymbol{x}_n, \boldsymbol{\theta}) = \prod_{n=1}^{N} \mathcal{N}\big(y_n \,|\, \boldsymbol{x}_n^\top \boldsymbol{\theta},\, \sigma^2\big)$$

minimize the negative log-likelihood

$$-\log p(\mathcal{Y} \,|\, \mathcal{X}, \boldsymbol{\theta}) = -\log \prod_{n=1}^{N} p(y_n \,|\, \boldsymbol{x}_n, \boldsymbol{\theta}) = -\sum_{n=1}^{N} \log p(y_n \,|\, \boldsymbol{x}_n, \boldsymbol{\theta})$$

LINEAR REGRESSION

Maximum Likelihood Estimation

minimize the negative log-likelihood

$$-\log p(\mathcal{Y} \mid \mathcal{X}, \boldsymbol{\theta}) = -\log \prod_{n=1}^{N} p(y_n \mid \boldsymbol{x}_n, \boldsymbol{\theta}) = -\sum_{n=1}^{N} \log p(y_n \mid \boldsymbol{x}_n, \boldsymbol{\theta})$$

$$\log p(y_n \mid \boldsymbol{x}_n, \boldsymbol{\theta}) = -\frac{1}{2\sigma^2}(y_n - \boldsymbol{x}_n^\top \boldsymbol{\theta})^2 + \text{const}$$

$$\text{log-likelihood } \mathcal{L}(\boldsymbol{\theta}) := \frac{1}{2\sigma^2}\sum_{n=1}^{N}(y_n - \boldsymbol{x}_n^\top \boldsymbol{\theta})^2 + \text{const} \; = \frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta})^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}) + \text{const}$$

$$= \frac{1}{2\sigma^2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}\|^2 + \text{const}$$

LINEAR REGRESSION

Maximum Likelihood Estimation

$$\text{log-likelihood } \mathcal{L}(\boldsymbol{\theta}) = \frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta})^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}) + \text{const}$$

$$\begin{aligned}
\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\boldsymbol{\theta}} &= \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}}\left(\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta})^\top(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\theta}) + \text{const}\right) \\
&= \frac{1}{2\sigma^2}\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}}\left(\boldsymbol{y}^\top\boldsymbol{y} - 2\boldsymbol{y}^\top\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\theta}^\top\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\theta}\right) \\
&= \frac{1}{\sigma^2}(-\boldsymbol{y}^\top\boldsymbol{X} + \boldsymbol{\theta}^\top\boldsymbol{X}^\top\boldsymbol{X}) \in \mathbb{R}^{1\times D}
\end{aligned}$$

LINEAR REGRESSION

Maximum Likelihood Estimation

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\boldsymbol{\theta}} = \frac{1}{\sigma^2}(-\boldsymbol{y}^\top \boldsymbol{X} + \boldsymbol{\theta}^\top \boldsymbol{X}^\top \boldsymbol{X}) \in \mathbb{R}^{1\times D}$$

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\boldsymbol{\theta}} = \boldsymbol{0}^\top$$

$$\boldsymbol{\theta}_{\mathrm{ML}}^\top \boldsymbol{X}^\top \boldsymbol{X} = \boldsymbol{y}^\top \boldsymbol{X}$$

$$\boldsymbol{\theta}_{\mathrm{ML}}^\top = \boldsymbol{y}^\top \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}$$

$$\boldsymbol{\theta}_{\mathrm{ML}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{y}$$

$\boldsymbol{X} := [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N]^\top \in \mathbb{R}^{N\times D}$ $\quad$ $\boldsymbol{y} := [y_1, \ldots, y_N]^\top \in \mathbb{R}^N$

LINEAR REGRESSION

Maximum Likelihood Estimation

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\boldsymbol{\theta}} = \frac{1}{\sigma^2}(-\boldsymbol{y}^\top \boldsymbol{X} + \boldsymbol{\theta}^\top \boldsymbol{X}^\top \boldsymbol{X}) \in \mathbb{R}^{1\times D}$$

$$\nabla^2_{\boldsymbol{\theta}}\mathcal{L}(\boldsymbol{\theta}) = \frac{1}{\sigma^2}\boldsymbol{X}^\top \boldsymbol{X} \text{ is positive definite.}$$

Definition 8.5 Positive Definite Matrices

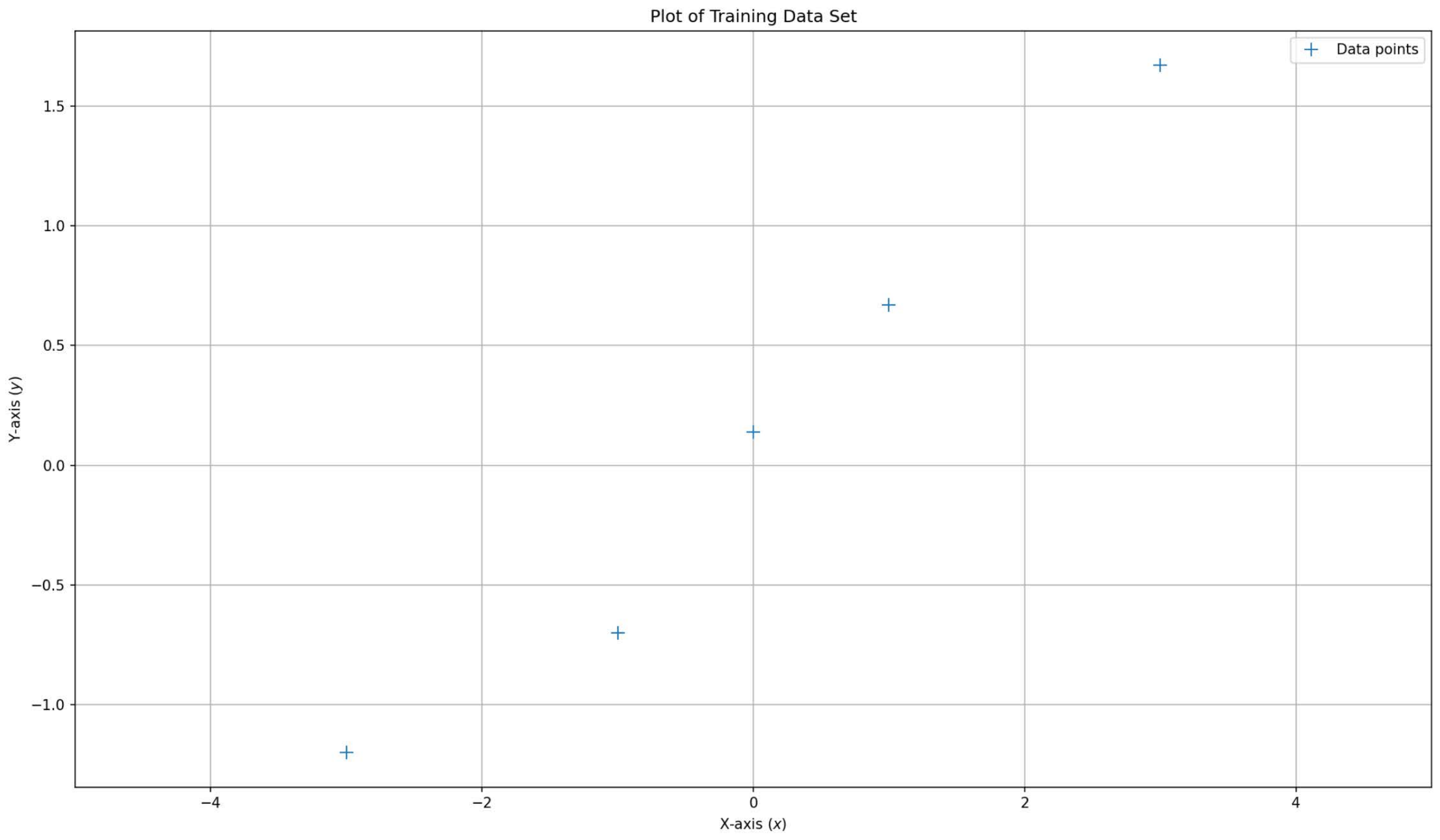
A square matrix is called ***positive definite*** *if it is symmetric and all its eigenvalues λ are positive, that is $\lambda > 0$.*

LINEAR REGRESSION

Maximum Likelihood Estimation

$$\theta_{\mathrm{ML}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$

```python
import numpy as np
import scipy.linalg
import matplotlib.pyplot as plt

# Define training set
X = np.array([-3, -1, 0.0, 1, 3]).reshape(-1,1) # 5x1 vector, N=5, D=1
y = np.array([-1.2, -0.7, 0.14, 0.67, 1.67]).reshape(-1,1) # 5x1 vector

# Plot the training set
plt.figure()
plt.plot(X, y, '+', markersize=10, label='Data points')
plt.xlabel("X-axis ($x$)")
plt.ylabel("Y-axis ($y$)")
plt.title("Plot of Training Data Set")
plt.xlim([-5, 5]) # Setting x-axis limits
plt.legend()
plt.grid(True)
plt.show()
```



LINEAR REGRESSION

Maximum Likelihood Estimation

$$\boldsymbol{\theta}_{\mathrm{ML}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$

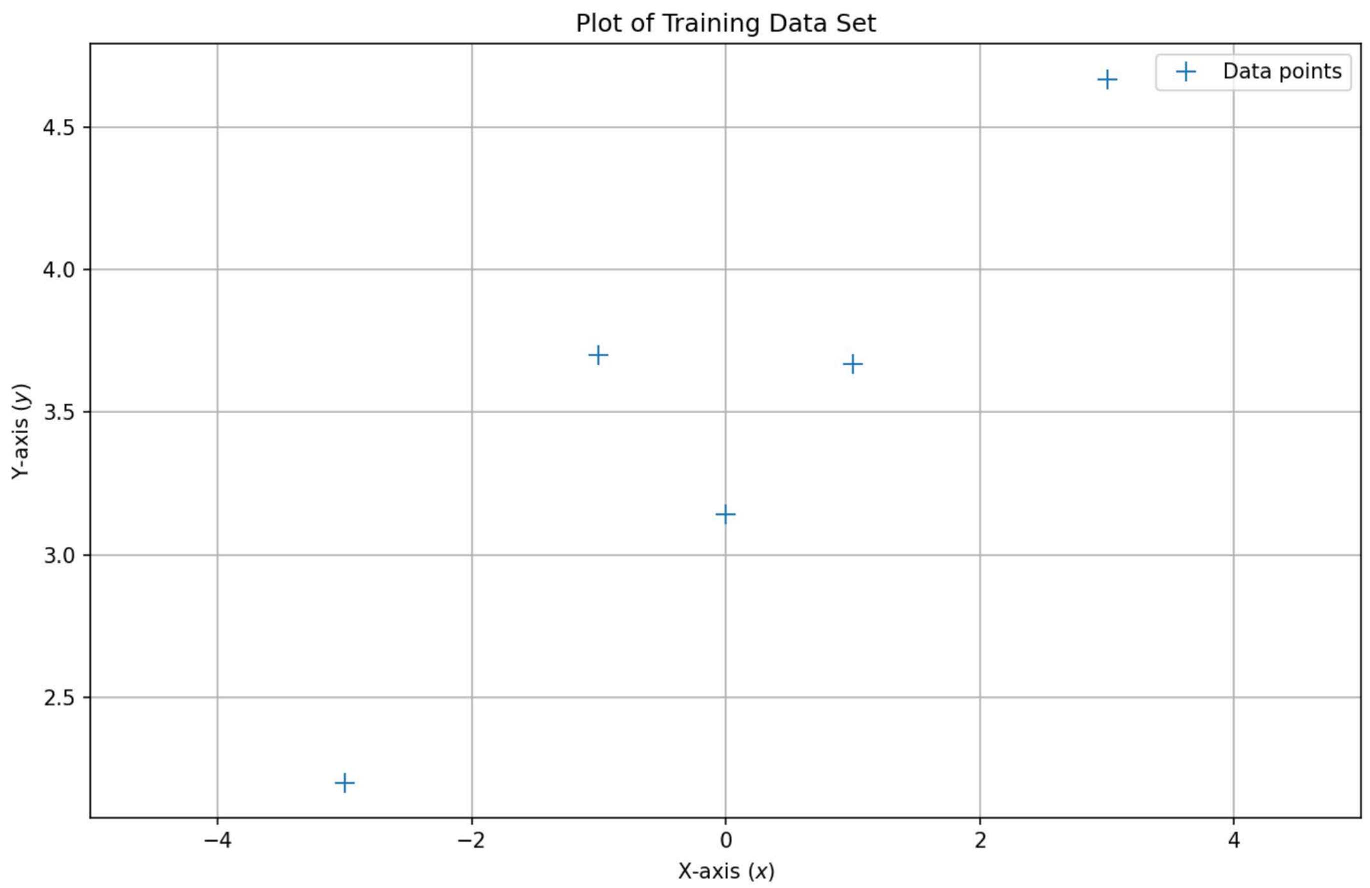
$$y = \Theta_{\mathrm{ml}} \times x$$

LINEAR REGRESSION

Maximum Likelihood Estimation

$$\boldsymbol{\theta}_{\mathrm{ML}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$

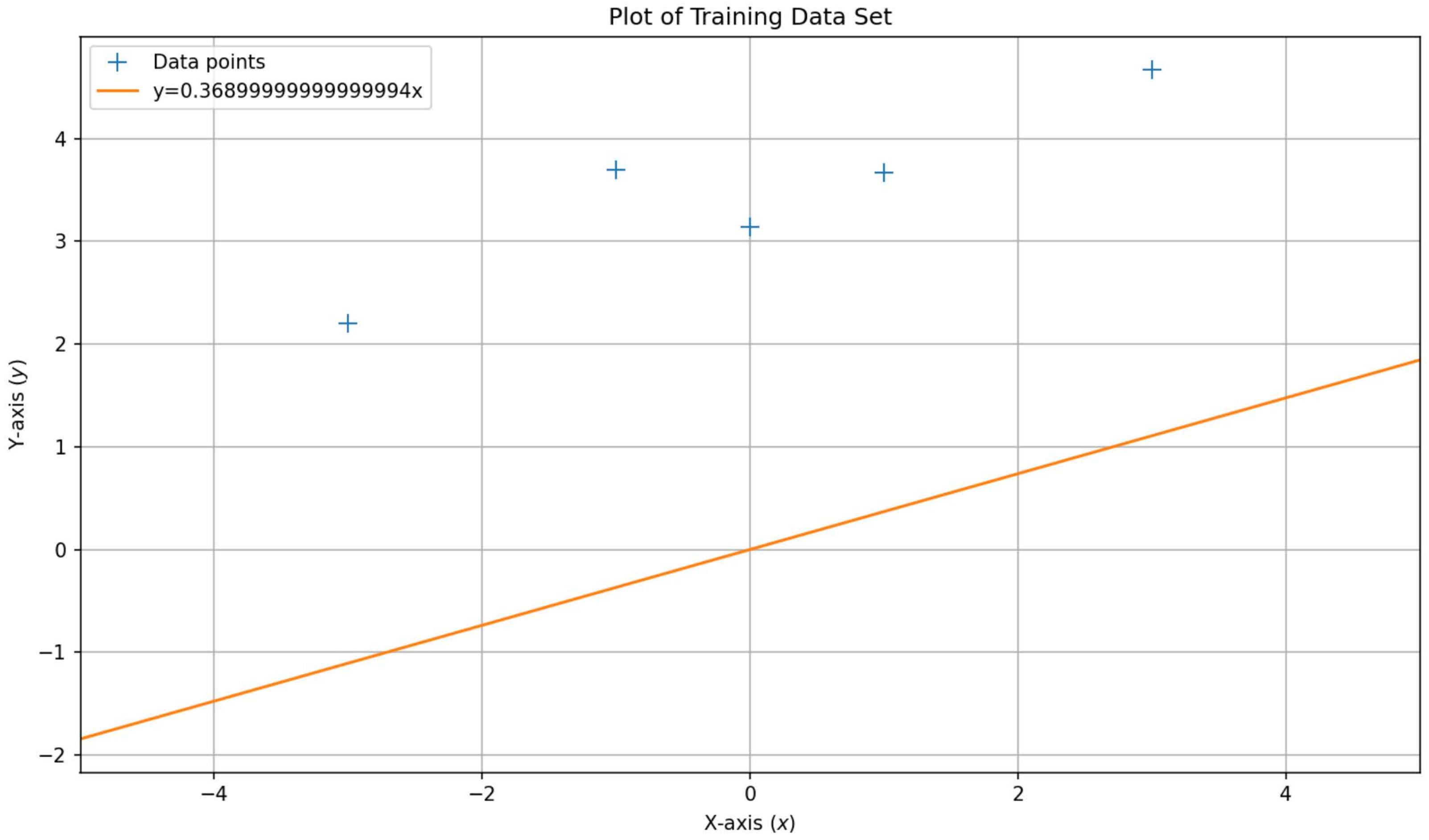
```python
import numpy as np
import scipy.linalg
import matplotlib.pyplot as plt

# Define training set
X = np.array([-3, -1, 0.0, 1, 3]).reshape(-1,1) # 5x1 vector, N=5, D=1
y = np.array([2.2, 3.7, 3.14, 3.67, 4.67]).reshape(-1,1) # 5x1 vector

# Plot the training set
plt.figure()
plt.plot(X, y, '+', markersize=10, label='Data points')
plt.xlabel("X-axis ($x$)")
plt.ylabel("Y-axis ($y$)")
plt.title("Plot of Training Data Set")
plt.xlim([-5, 5]) # Setting x-axis limits
plt.legend()
plt.grid(True)
plt.show()
```



LINEAR REGRESSION

Maximum Likelihood Estimation

$$\boldsymbol{\theta}_{\mathrm{ML}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$



LINEAR REGRESSION

Maximum Likelihood Estimation

$$\boldsymbol{\theta}_{\mathrm{ML}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$

$$\boxed{y = \Theta_{\mathrm{ml}} \times x}$$

Slope-Intercept Form:

$$y = mx + b$$

m is the slope of the line (how steep it is),

b is the y-intercept (where the line crosses the y-axis).

LINEAR REGRESSION

Maximum Likelihood Estimation

$$\boldsymbol{\theta}_{\mathrm{ML}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$

Hint: augmented training inputs of size N x (D+1)

CODING TIME !!!!

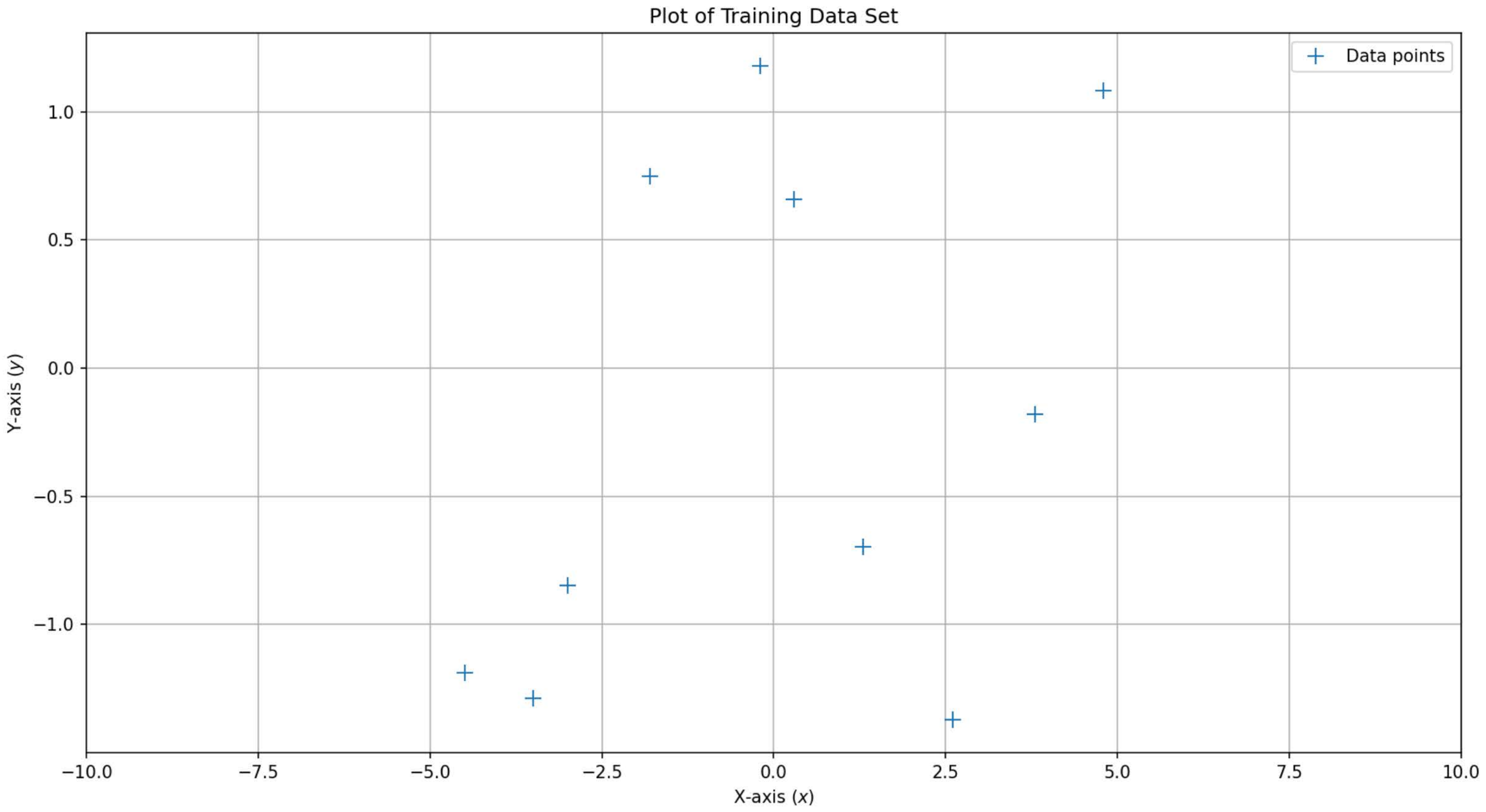
MAXIMUM LIKELIHOOD ESTIMATION WITH FEATURES

LINEAR REGRESSION

Maximum Likelihood Estimation with Features

$$\theta_{\text{ML}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$ ❓

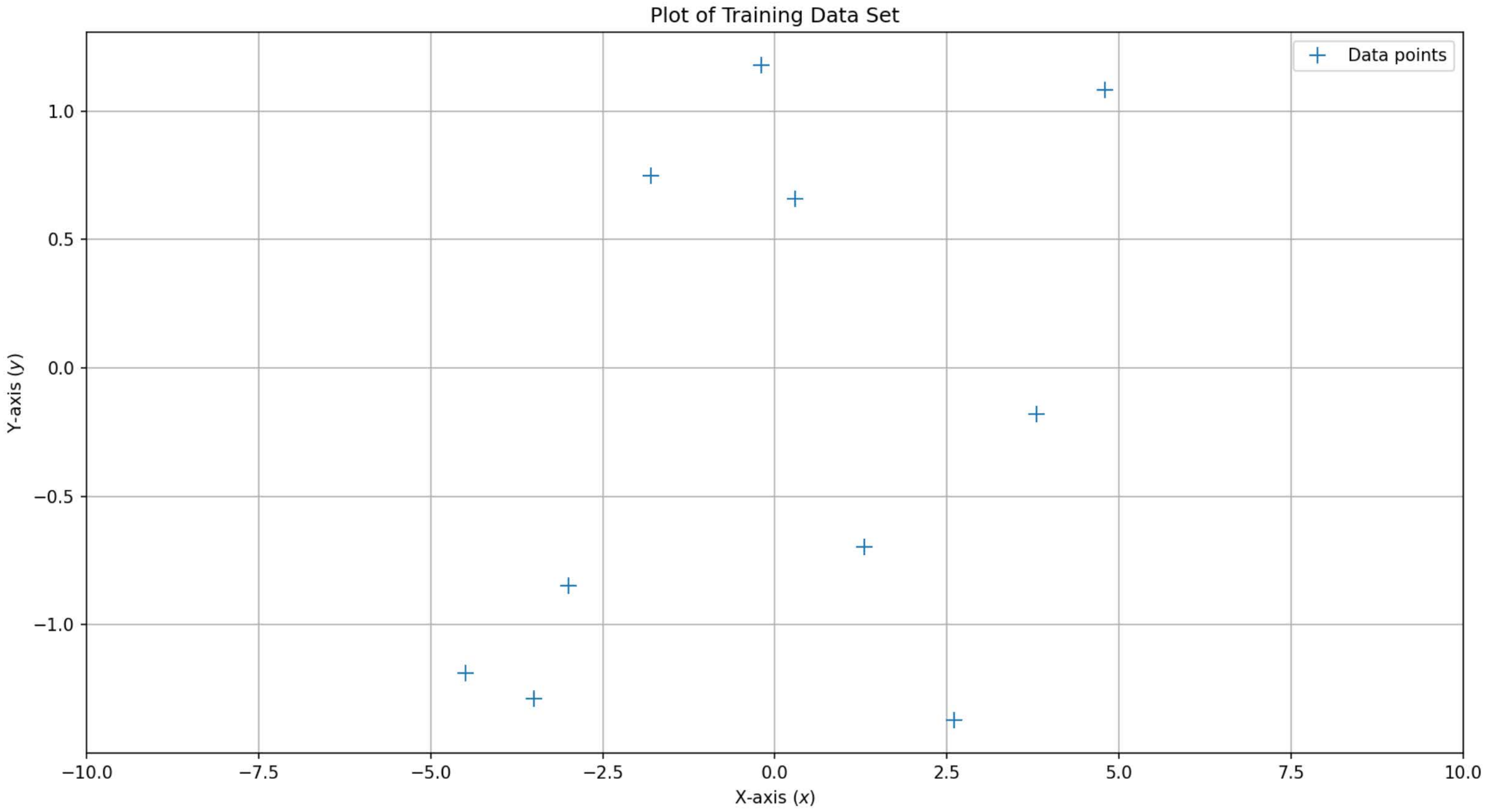
```python
import numpy as np
import scipy.linalg
import matplotlib.pyplot as plt

X = np.array([-4.5, -3.5, -3, -1.8, -0.2, 0.3, 1.3, 2.6, 3.8, 4.8]).reshape(-1,1) # 5x1 vector, N=5, D=1
y = np.array([
    [-1.11362822],
    [-1.24394281],
    [-0.91157385],
    [0.67067171],
    [1.24891634],
    [0.7776148],
    [-0.62067303],
    [-1.41641754],
    [-0.30383694],
    [0.92323755]
]).reshape(-1,1) # 5x1 vector
```

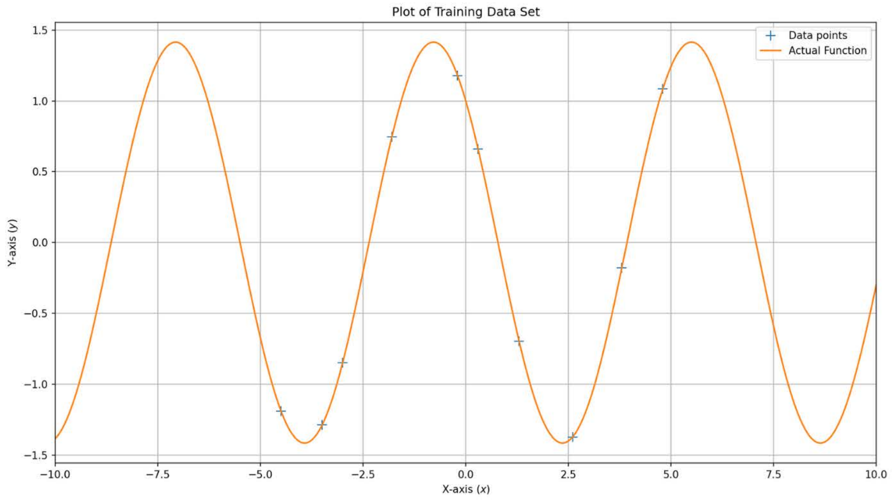


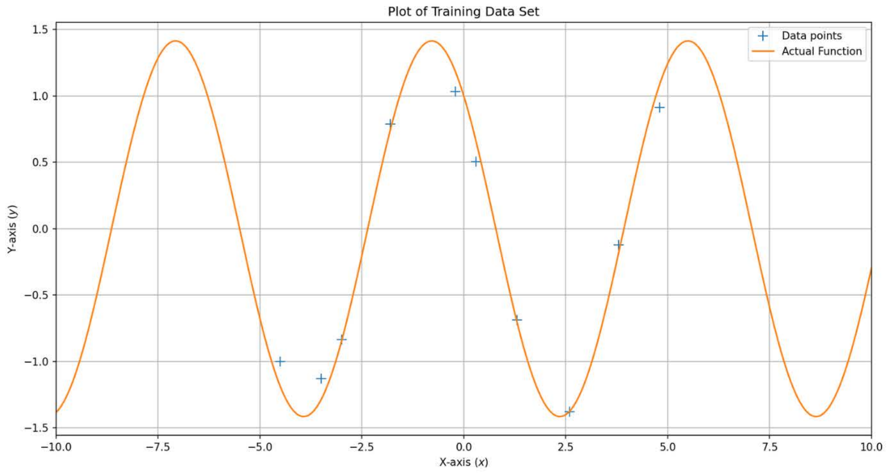
LINEAR REGRESSION

Maximum Likelihood Estimation with Features

$$\theta_{\mathrm{ML}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$



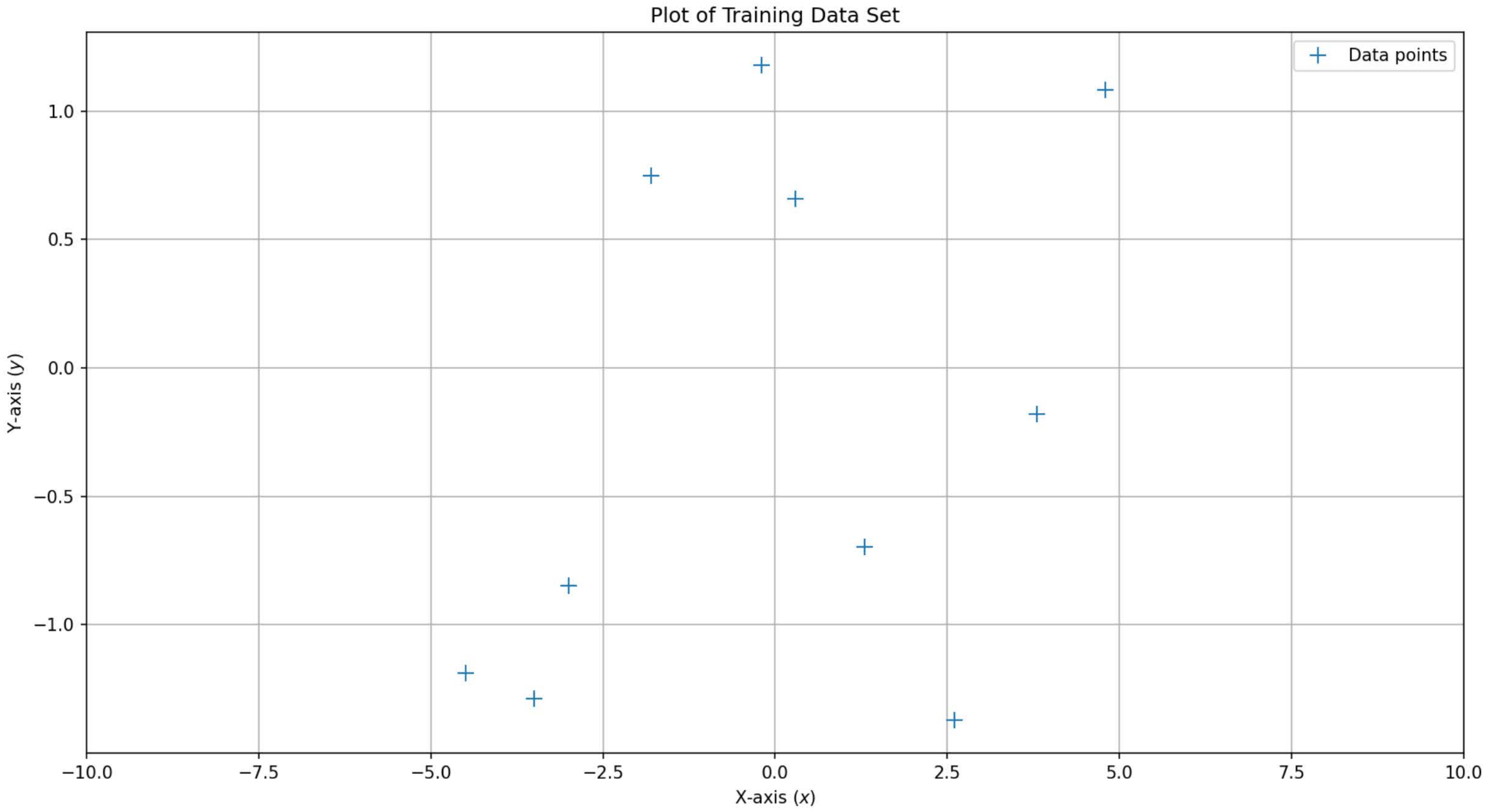


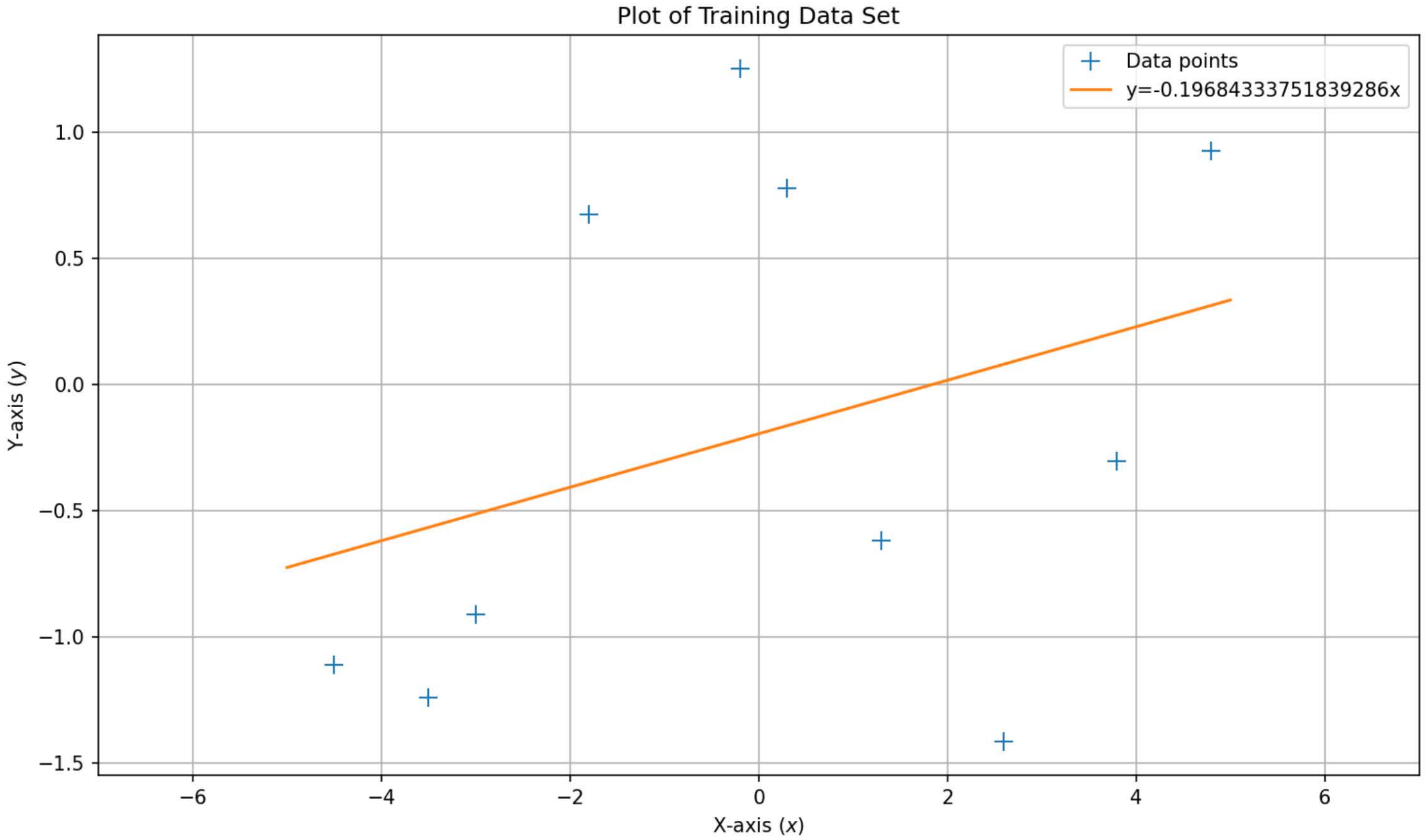


LINEAR REGRESSION

Maximum Likelihood Estimation with Features

$$\theta_{\mathrm{ML}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$ ❓





LINEAR REGRESSION

Maximum Likelihood Estimation with Features

$$p(y \mid \boldsymbol{x}, \boldsymbol{\theta}) = \mathcal{N}\big(y \mid \boldsymbol{\phi}^\top(\boldsymbol{x})\boldsymbol{\theta},\, \sigma^2\big) \iff y = \boldsymbol{\phi}^\top(\boldsymbol{x})\boldsymbol{\theta} + \epsilon = \sum_{k=0}^{K-1} \theta_k \phi_k(\boldsymbol{x}) + \epsilon$$

$$\boldsymbol{\Phi} := \begin{bmatrix} \boldsymbol{\phi}^\top(\boldsymbol{x}_1) \\ \vdots \\ \boldsymbol{\phi}^\top(\boldsymbol{x}_N) \end{bmatrix} = \begin{bmatrix} \phi_0(\boldsymbol{x}_1) & \cdots & \phi_{K-1}(\boldsymbol{x}_1) \\ \phi_0(\boldsymbol{x}_2) & \cdots & \phi_{K-1}(\boldsymbol{x}_2) \\ \vdots & & \vdots \\ \phi_0(\boldsymbol{x}_N) & \cdots & \phi_{K-1}(\boldsymbol{x}_N) \end{bmatrix} \in \mathbb{R}^{N \times K}$$

LINEAR REGRESSION

Maximum Likelihood Estimation with Features

(Polynomial Regression)

$$\phi(x) = \begin{bmatrix} \phi_0(x) \\ \phi_1(x) \\ \vdots \\ \phi_{K-1}(x) \end{bmatrix} = \begin{bmatrix} 1 \\ x \\ x^2 \\ x^3 \\ \vdots \\ x^{K-1} \end{bmatrix}$$

(Feature Matrix for Second-order Polynomials)

$$\boldsymbol{\Phi} = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_N & x_N^2 \end{bmatrix}$$

LINEAR REGRESSION

Maximum Likelihood Estimation with Features

$$p(y \,|\, \boldsymbol{x}, \boldsymbol{\theta}) = \mathcal{N}\left(y \,|\, \boldsymbol{\phi}^\top(\boldsymbol{x})\boldsymbol{\theta},\, \sigma^2\right) \iff y = \boldsymbol{\phi}^\top(\boldsymbol{x})\boldsymbol{\theta} + \epsilon = \sum_{k=0}^{K-1} \theta_k \phi_k(\boldsymbol{x}) + \epsilon$$

$$-\log p(\mathcal{Y} \,|\, \mathcal{X}, \boldsymbol{\theta}) = \frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\theta})^\top(\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\theta}) + \text{const}\,.$$

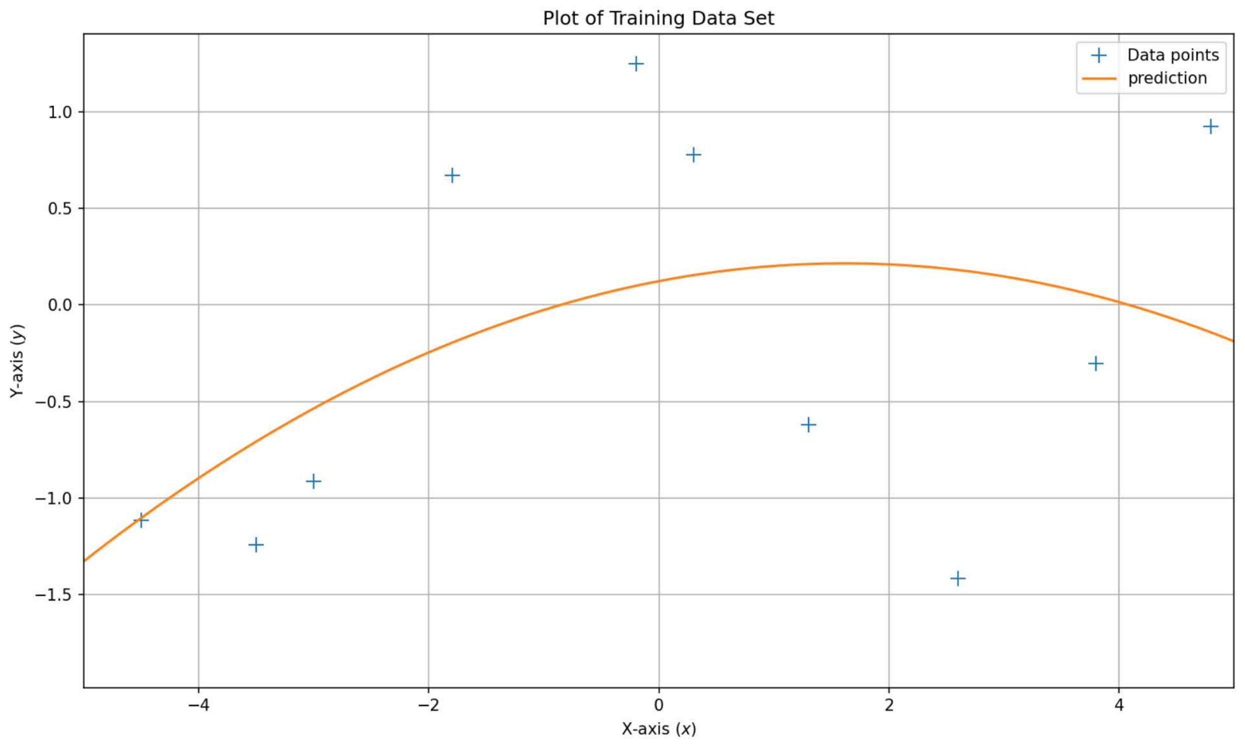
$$\boldsymbol{\theta}_{\text{ML}} = (\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top\boldsymbol{y}$$

LINEAR REGRESSION

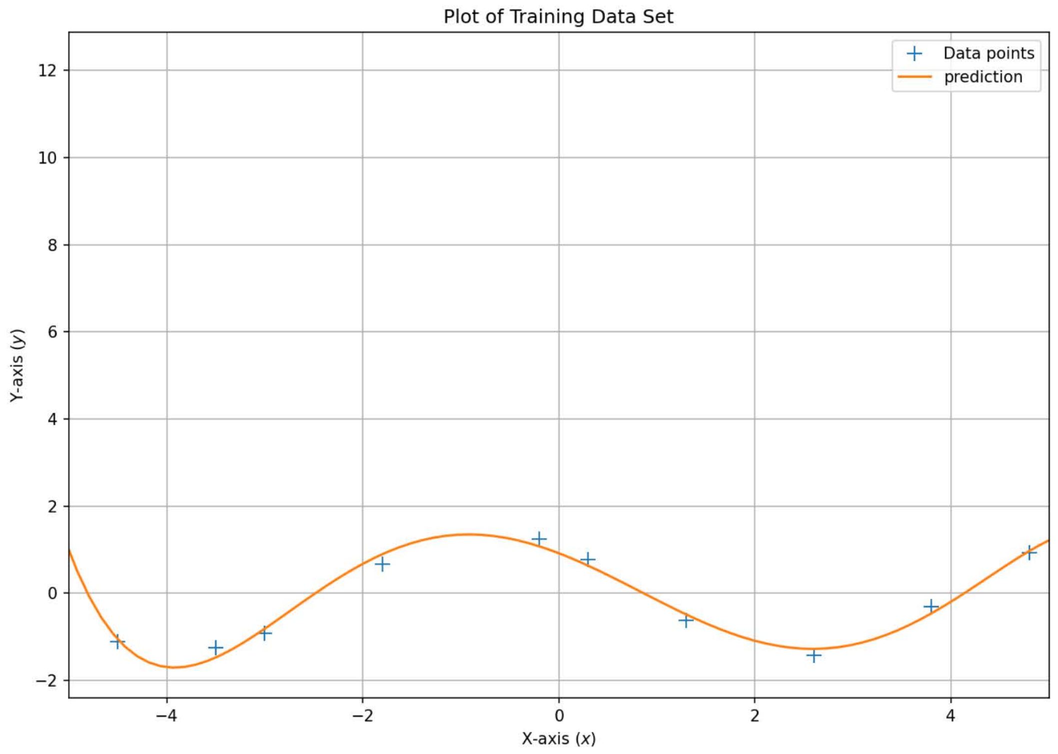
Maximum Likelihood Estimation with Features

$$\theta_{\mathrm{ML}} = (\Phi^\top \Phi)^{-1} \Phi^\top y$$

K=2



K=5



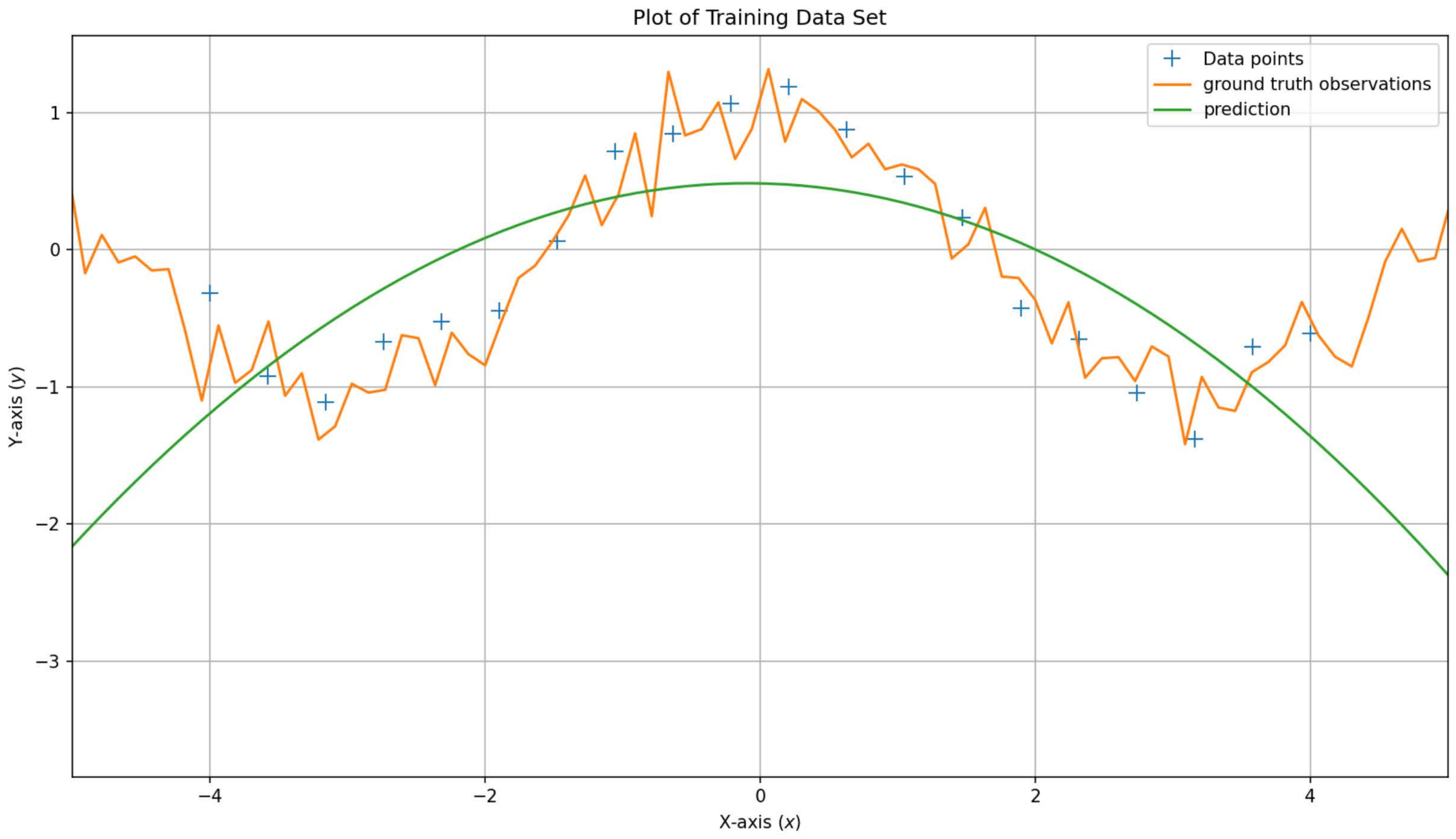
LINEAR REGRESSION

Maximum Likelihood Estimation with Features

$$\theta_{\mathrm{ML}} = (\Phi^\top \Phi)^{-1} \Phi^\top y$$

K=2

```python
import numpy as np
import scipy.linalg
import matplotlib.pyplot as plt

def f(x):
    return np.cos(x) + 0.2*np.random.normal(size=(x.shape))

X = np.linspace(-4,4,20).reshape(-1,1)
y = f(X)
```



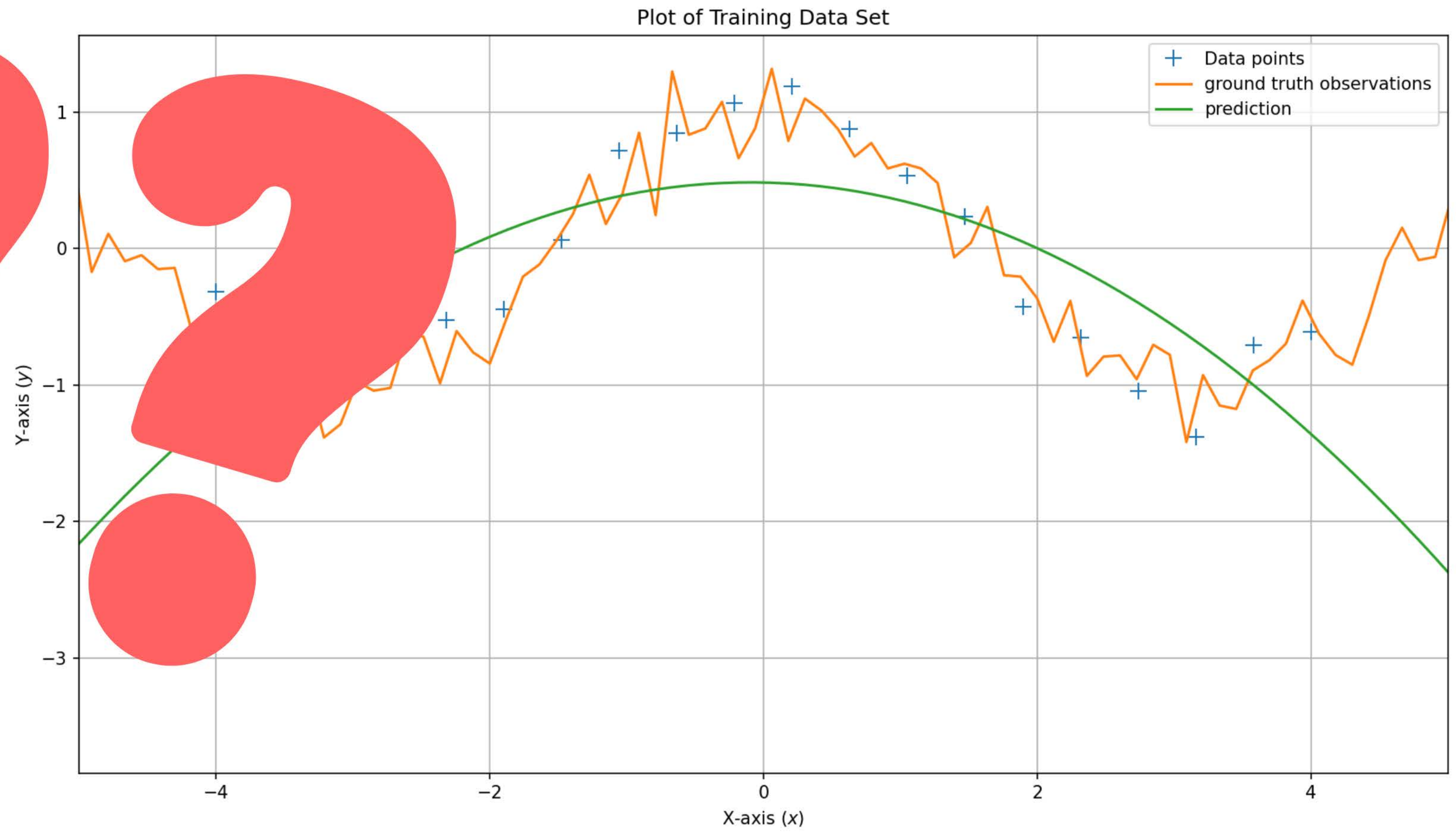
LINEAR REGRESSION

Maximum Likelihood Estimation with Features

$$\theta_{\mathrm{ML}} = (\Phi^\top \Phi)^{-1} \Phi^\top y$$

K=2

```python
import numpy as np
import scipy.linalg
import matplotlib.pyplot as plt

def f(x):
    return np.cos(x) + 0.2*np.random.normal(size=(x.shape))

X = np.linspace(-4,4,20).reshape(-1,1)
y = f(X)
```

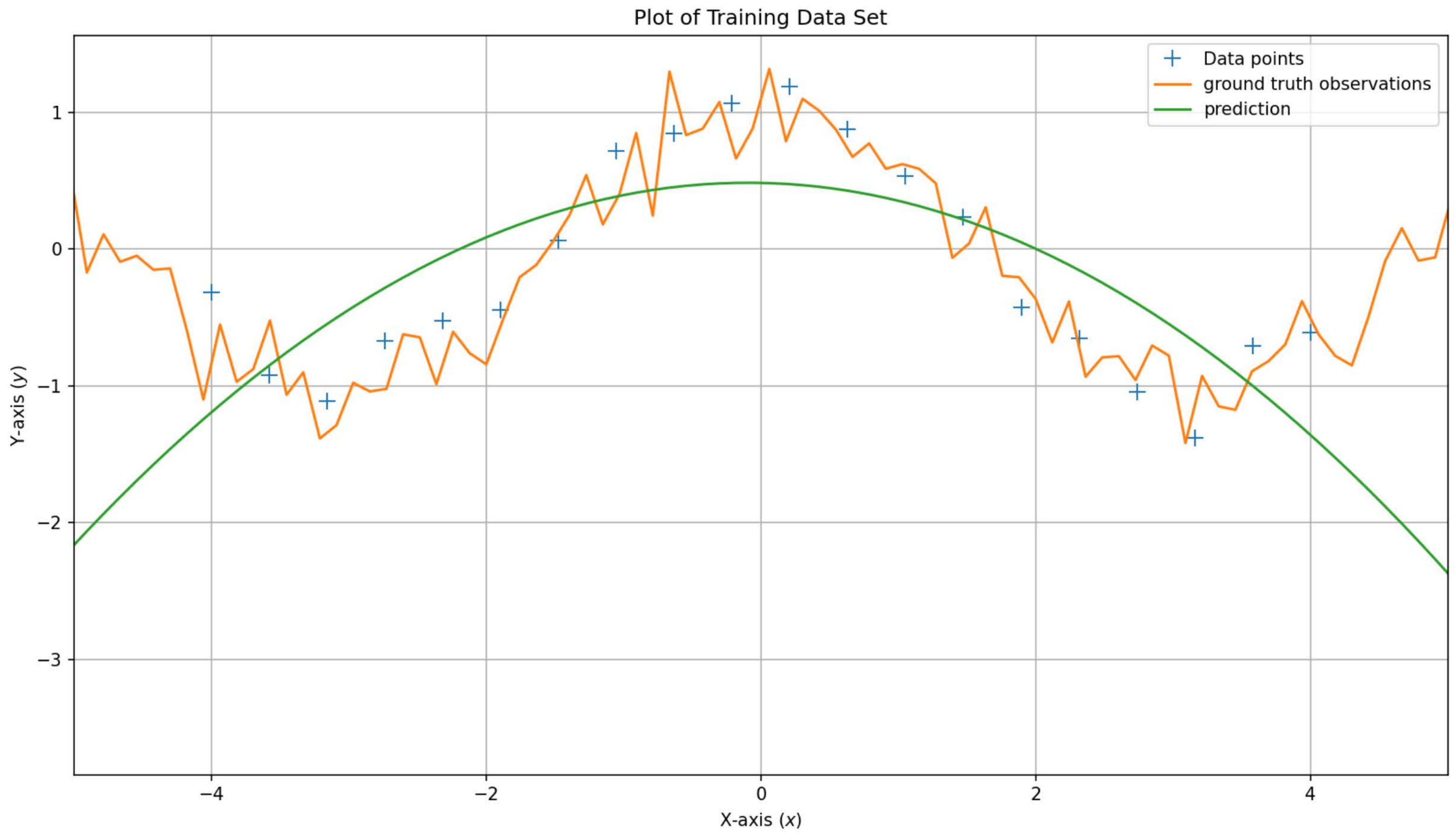


OVERFITTING IN LINEAR REGRESSION

LINEAR REGRESSION

Overfitting in Linear Regression

K=2



Loss Function

root mean square error (RMSE)

$$\sqrt{\frac{1}{N}\|\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{\theta}\|^2} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n - \boldsymbol{\phi}^\top(\boldsymbol{x}_n)\boldsymbol{\theta})^2}$$

LINEAR REGRESSION

Overfitting in Linear Regression

Loss Function

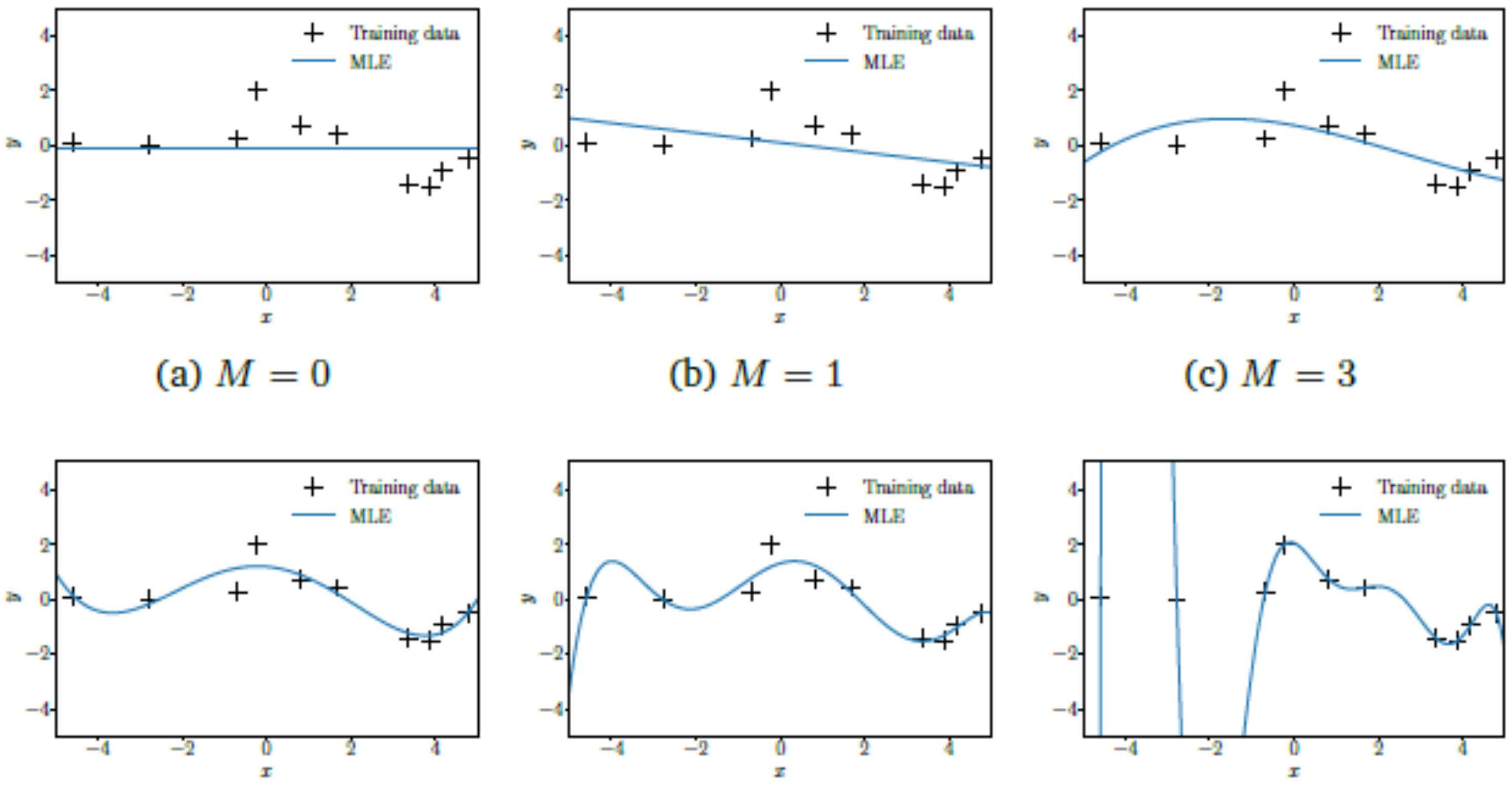
root mean square error (RMSE)

$$\sqrt{\frac{1}{N}\|\boldsymbol{y}-\boldsymbol{\Phi}\boldsymbol{\theta}\|^2}=\sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_n-\boldsymbol{\phi}^\top(\boldsymbol{x}_n)\boldsymbol{\theta})^2}$$

```
def RMSE(y, ypred):
    rmse = np.sqrt(np.mean((y-ypred)**2))
    return rmse
```

LINEAR REGRESSION

Overfitting in Linear Regression

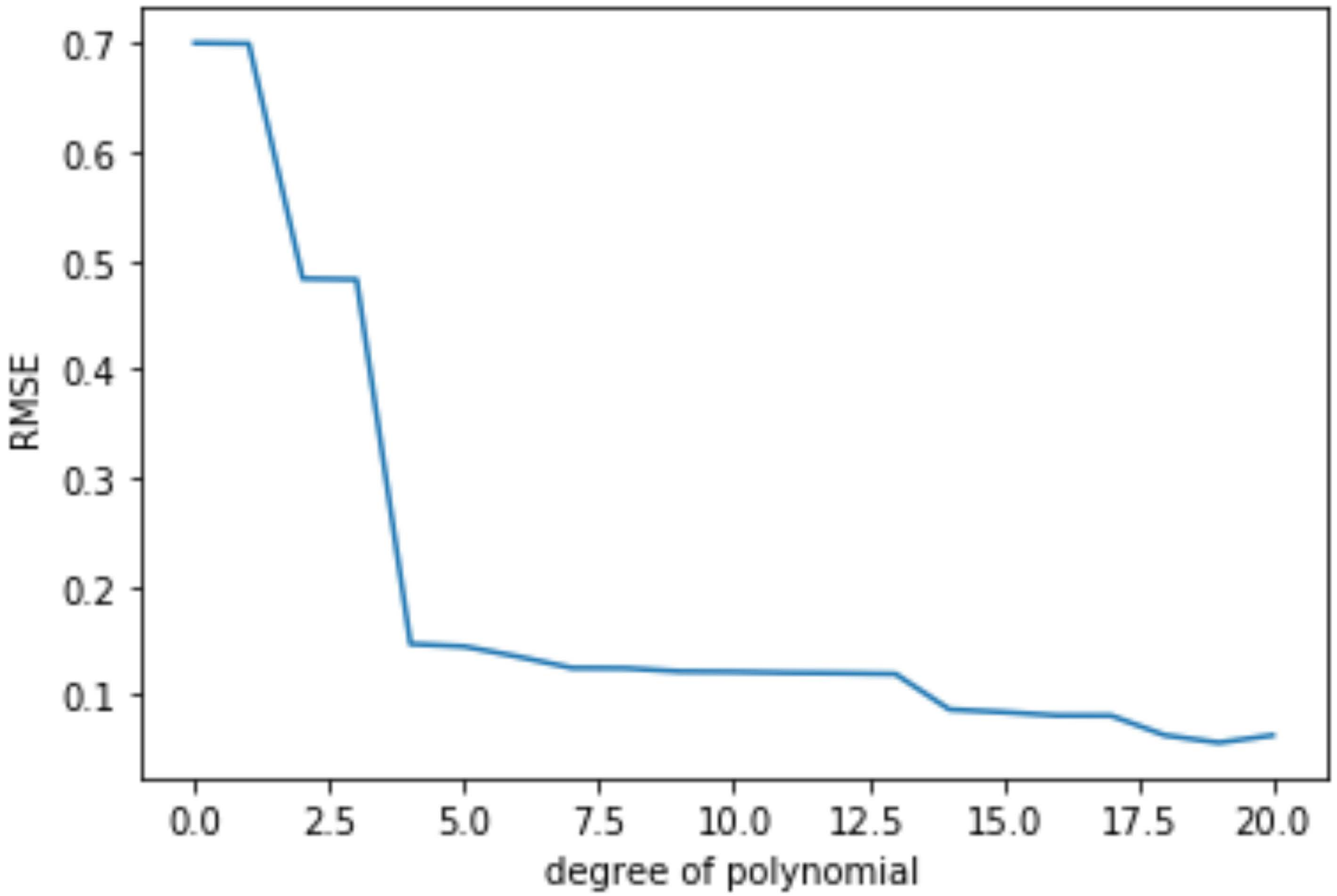


(a) $M = 0$ (b) $M = 1$ (c) $M = 3$

(d) $M = 4$ (e) $M = 6$ (f) $M = 9$

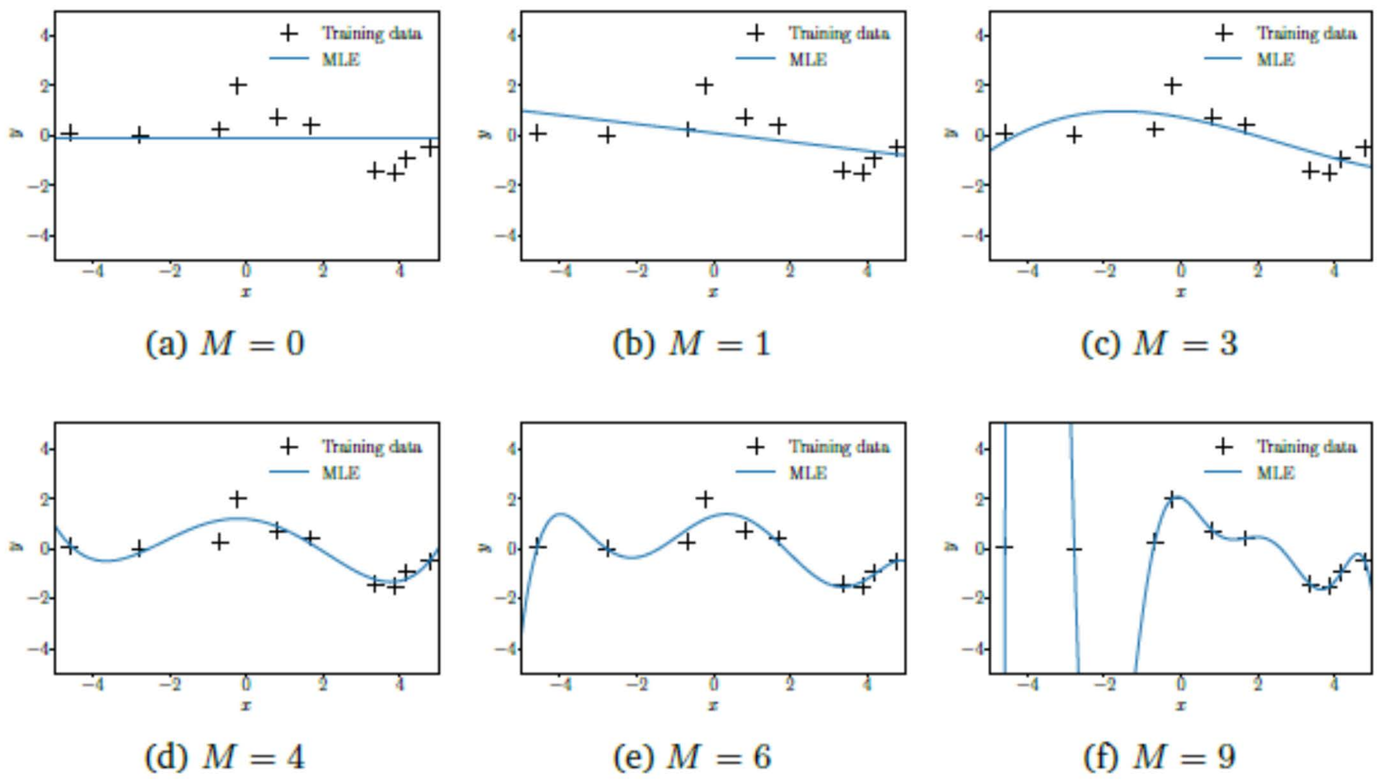
LINEAR REGRESSION

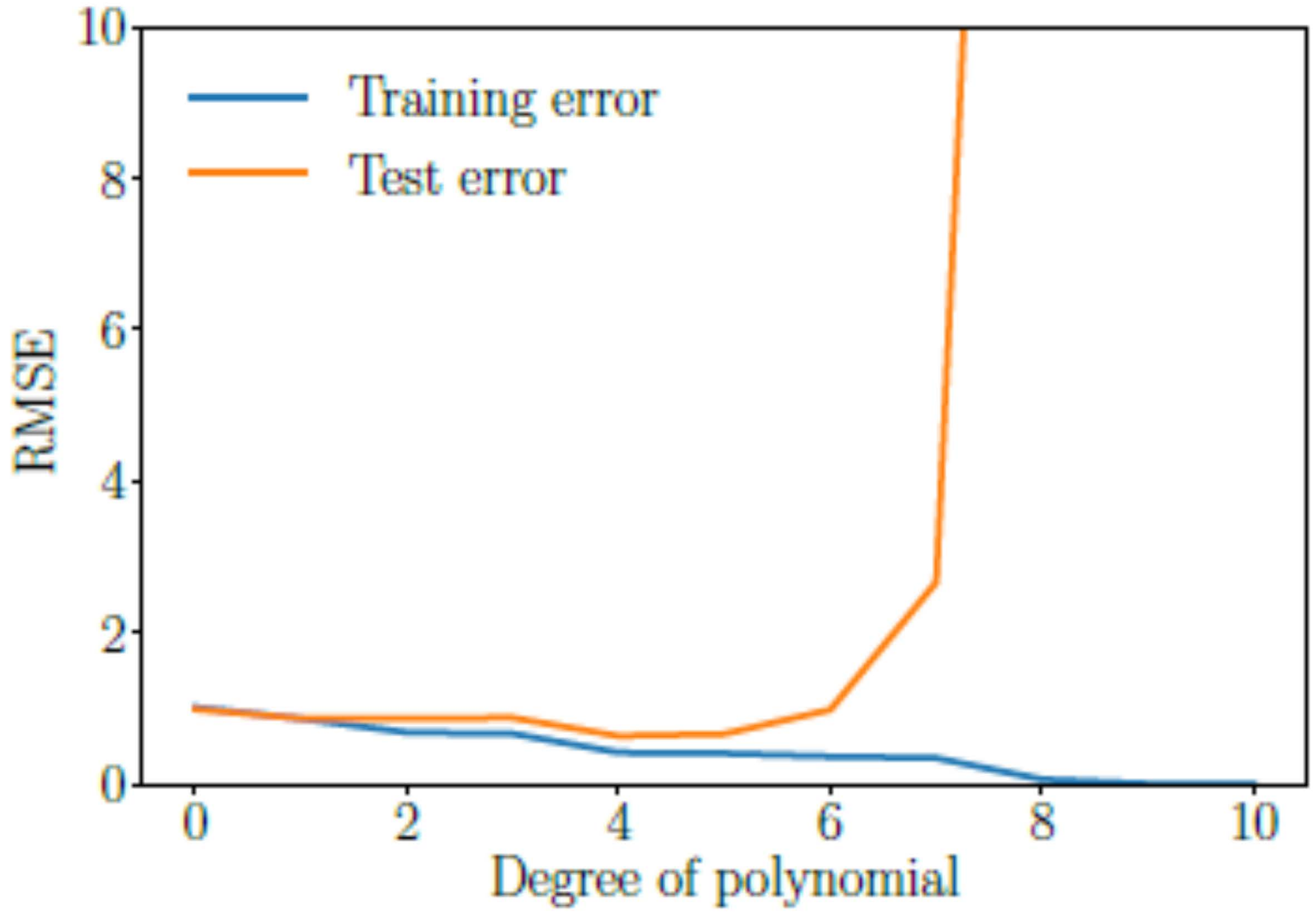
Overfitting in Linear Regression



LINEAR REGRESSION

Overfitting in Linear Regression





THANK YOU FOR LISTENING

Assignment 3

```python
import numpy as np
import scipy.linalg
import matplotlib.pyplot as plt

X = np.array([-4.5, -3.5, -3, -1.8, -0.2, 0.3, 1.3, 2.6, 3.8, 4.8]).reshape(-1,1) # 10x1 vector, N=5, D=1
y = np.array([
    [-0.91650116],
    [-0.47546053],
    [-0.10972425],
    [0.29504095],
    [-0.01596218],
    [0.10014949],
    [0.48104303],
    [0.10979023],
    [-0.99742128],
    [-0.91221826]
]).reshape(-1,1) # 5x1 vector

X_test = np.array([-3.99, -1.38, -1.37, -0.94,  0.69,  1.4,   1.57,  1.78,  1.81,  4.89]).reshape(-1,1) # 10x1 vector, N=5, D=1
y_test = np.array([
    [-0.80737607],
    [0.19813376],
    [0.19537639],
    [0.07185977],
    [0.24954213],
    [0.50662504],
    [0.52943298],
    [0.52406997],
    [0.51999057],
    [-0.82318288]
]).reshape(-1,1) # 10x1 vector, N=5, D=1
```