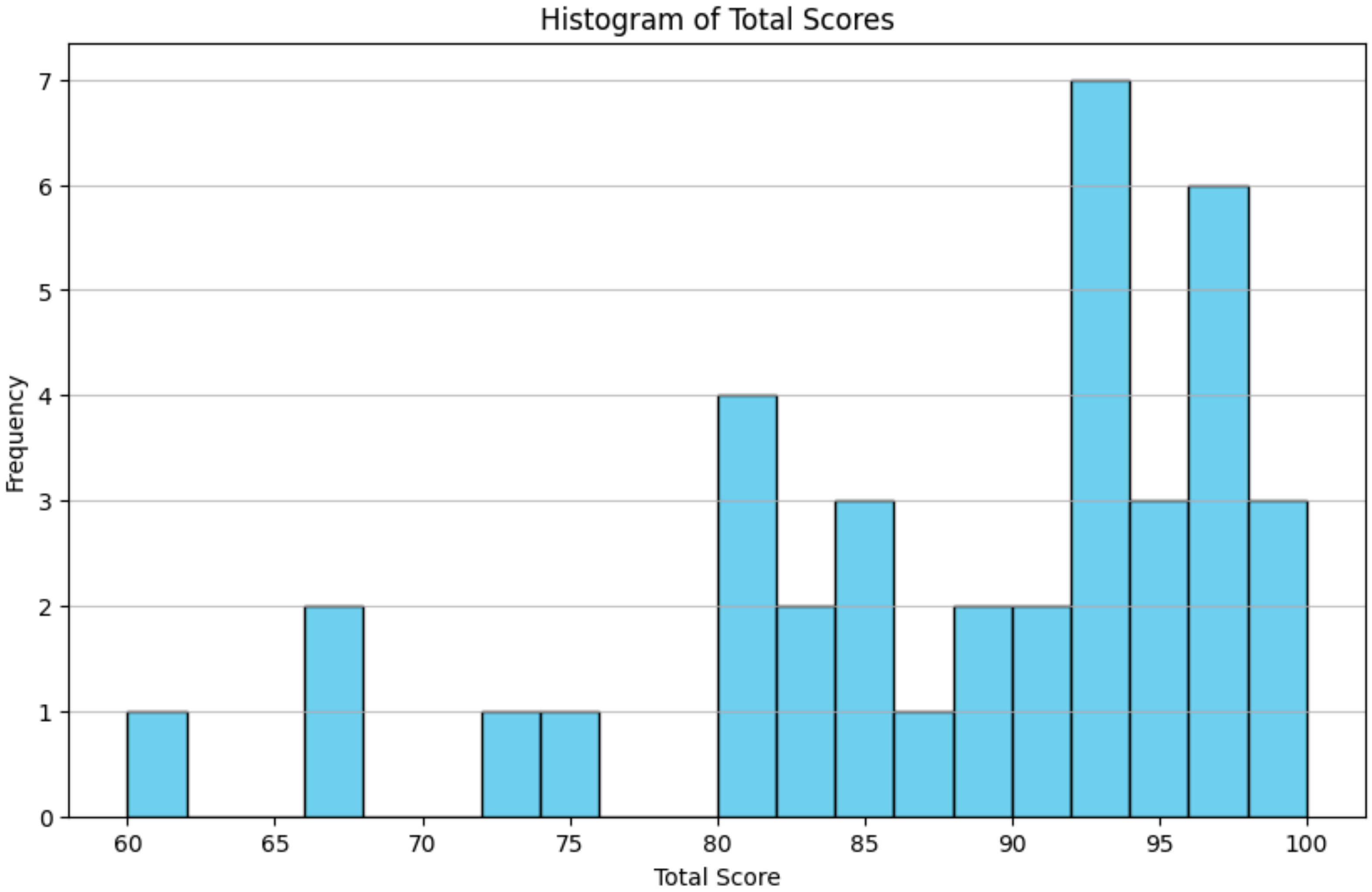
NYU

Lecture 13

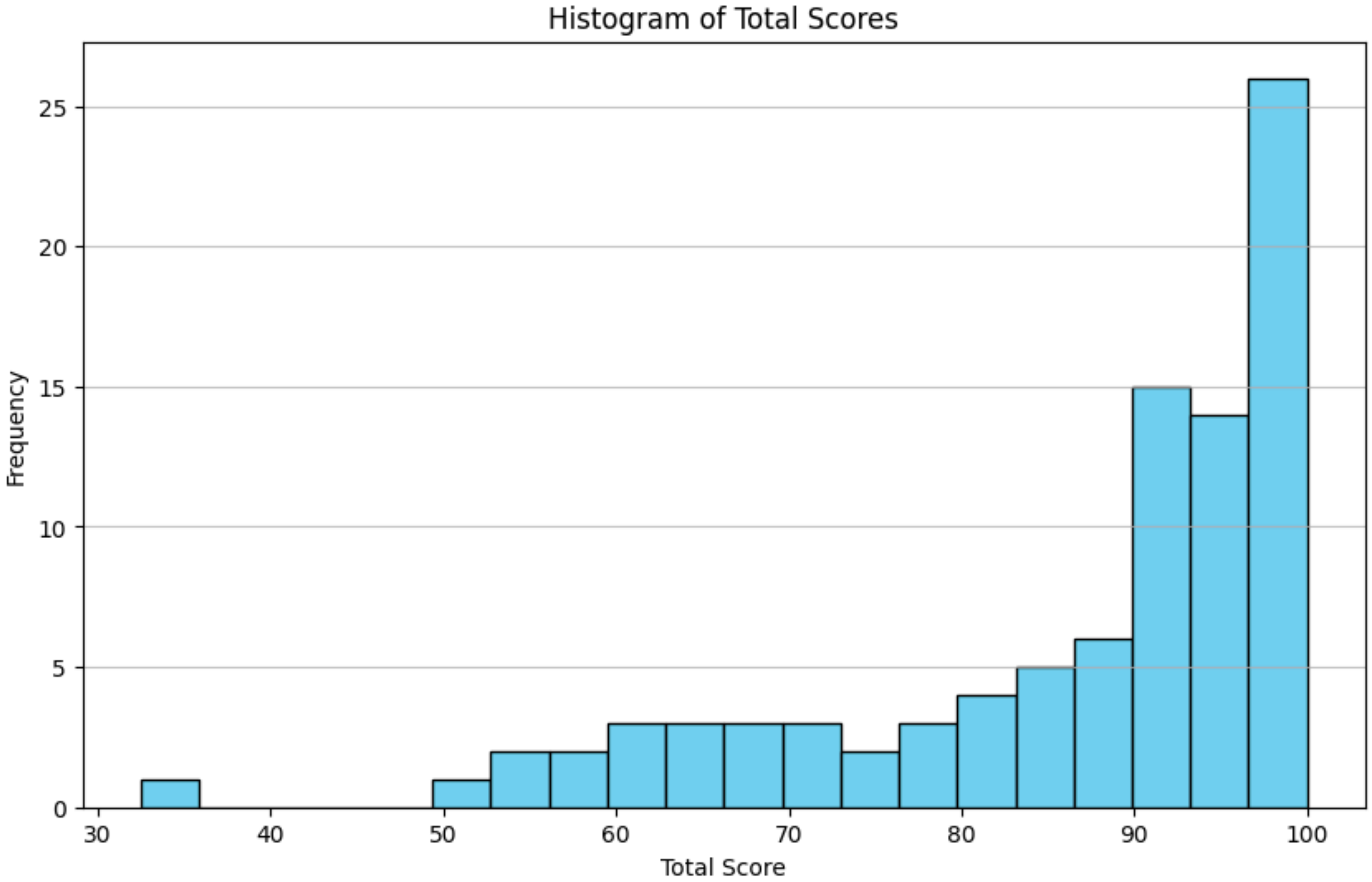
Least Squares

Dr. Yiping Lu

- Mean 87.95 median 91.75



- Mean 86.7 median 92.6



NYU

This Meeting is Being Recorded

Strang Sections 4.3 – Least Squares Approximations

Course notes adapted from N. Hammoud's NYU lecture notes.

Best-Fit Line

Example

Example: Find the best-fit line through the points $(0, 6)$, $(1, 0)$, and $(2, 0)$.

Example

Example: Find the best-fit line through the points $(0, 6)$, $(1, 0)$, and $(2, 0)$.

Example

Example: Find the best-fit line through the points $(0, 6)$, $(1, 0)$, and $(2, 0)$.

Least Squares Solution

Least Squares Solution

Let A be an $m \times n$ matrix.

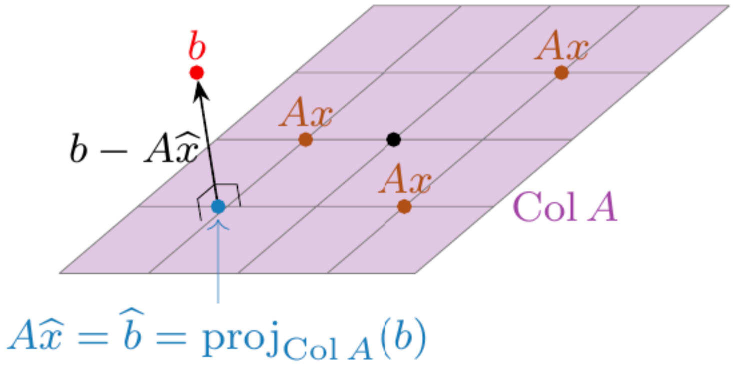
Definition

A **least squares solution** to $Ax = b$ is a vector $\widehat{x}$ in $\mathbf{R}^n$ such that

$$\|b - A\widehat{x}\| \leq \|b - Ax\|$$

for all x in $\mathbf{R}^n$.

Note that $b - A\widehat{x}$ is in $(\mathrm{Col}\, A)^{\perp}$.



In other words, a least squares solution $\widehat{x}$ solves $Ax = b$ *as closely as possible*.

Equivalently, a least squares solution to $Ax = b$ is a vector $\widehat{x}$ in $\mathbf{R}^n$ such that

$$A\widehat{x} = \widehat{b} = \mathrm{proj}_{\mathrm{Col}\, A}(b).$$

This is because $\widehat{b}$ is the closest vector to b such that $A\widehat{x} = \widehat{b}$ is consistent.

Least Squares Solution

Theorem

The least squares solutions to $Ax = b$ are the solutions to

$$(A^T A)\widehat{x} = A^T b.$$

This is just another $Ax = b$ problem, but with a *square* matrix $A^T A$!
Note we compute $\widehat{x}$ directly, without computing $\widehat{b}$ first.

Theorem

Let A be an $m \times n$ matrix. The following are equivalent:

1. $Ax = b$ has a *unique* least squares solution for all b in $\mathbf{R}^n$.
2. The columns of A are linearly independent.
3. $A^T A$ is invertible.

In this case, the least squares solution is $(A^T A)^{-1}(A^T b)$.

Least Squares Solution – Yesterday's Example

Find the least squares solutions to $Ax = b$ where:

$$A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \qquad b = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}.$$

We have

$$A^T A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & 3 \\ 3 & 5 \end{pmatrix}$$

and

$$A^T b = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 6 \\ 0 \end{pmatrix}.$$

Row reduce:

$$\left(\begin{array}{cc|c} 3 & 3 & 6 \\ 3 & 5 & 0 \end{array}\right) \rightsquigarrow \left(\begin{array}{cc|c} 1 & 0 & 5 \\ 0 & 1 & -3 \end{array}\right).$$

So the only least squares solution is $\widehat{x} = \begin{pmatrix} 5 \\ -3 \end{pmatrix}$.

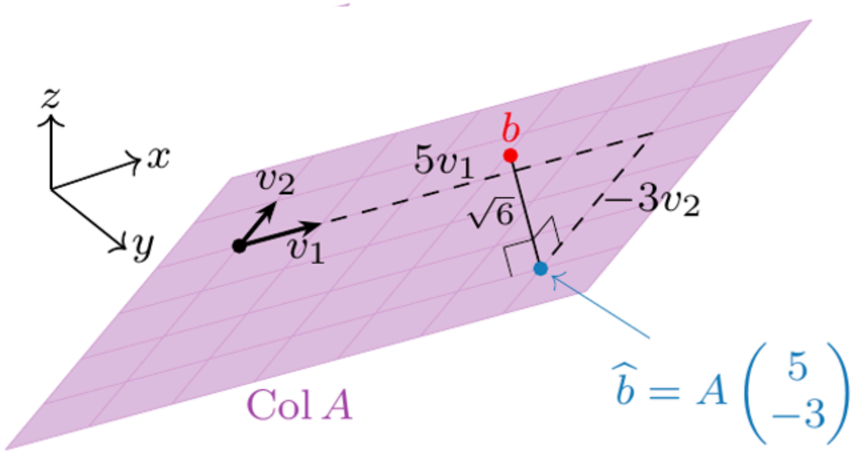
Least Squares Solution – Worked Example

How close did we get?

$$\widehat{b} = A\widehat{x} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 5 \\ -3 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \\ -1 \end{pmatrix}$$

The distance from b is

$$\|b - A\widehat{x}\| = \left\| \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 5 \\ 2 \\ -1 \end{pmatrix} \right\| = \left\| \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \right\| = \sqrt{1^2 + (-2)^2 + 1^2} = \sqrt{6}.$$



Least Squares Solution

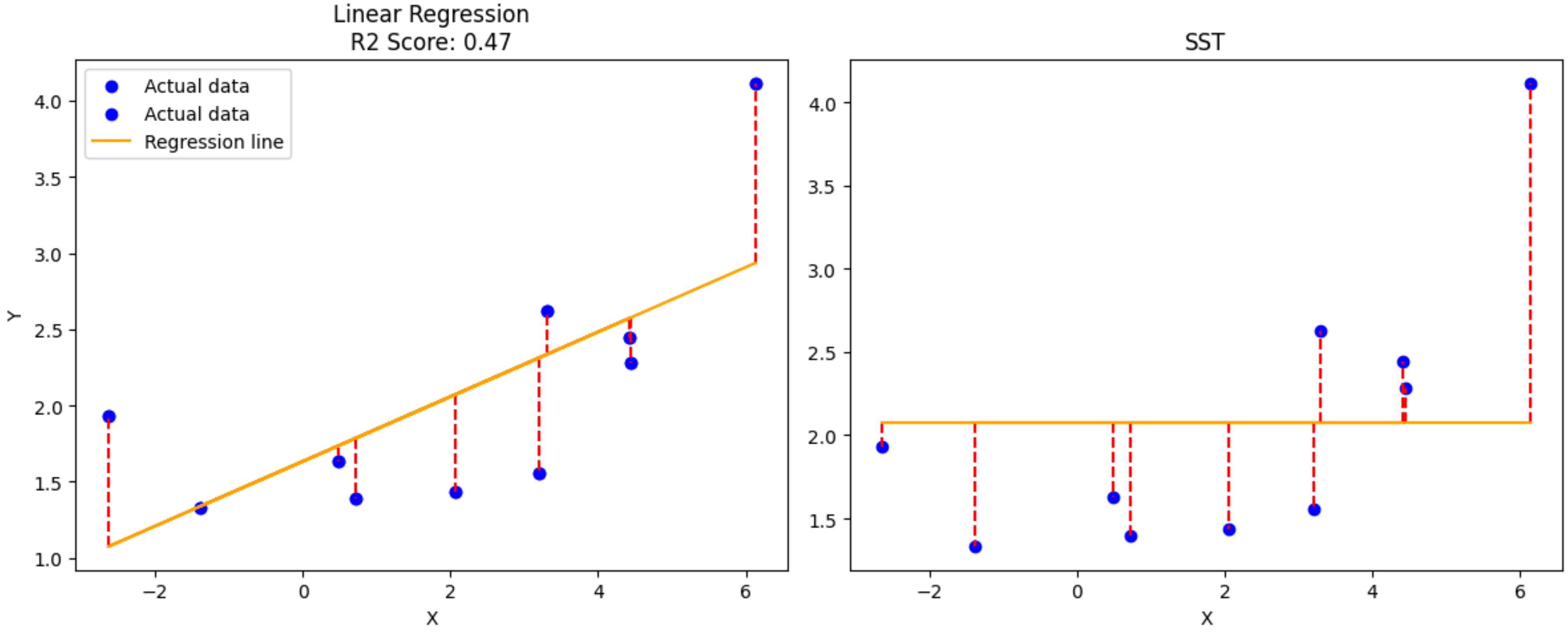
Example: Find the least squares solution to $\begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} \vec{x} = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$.

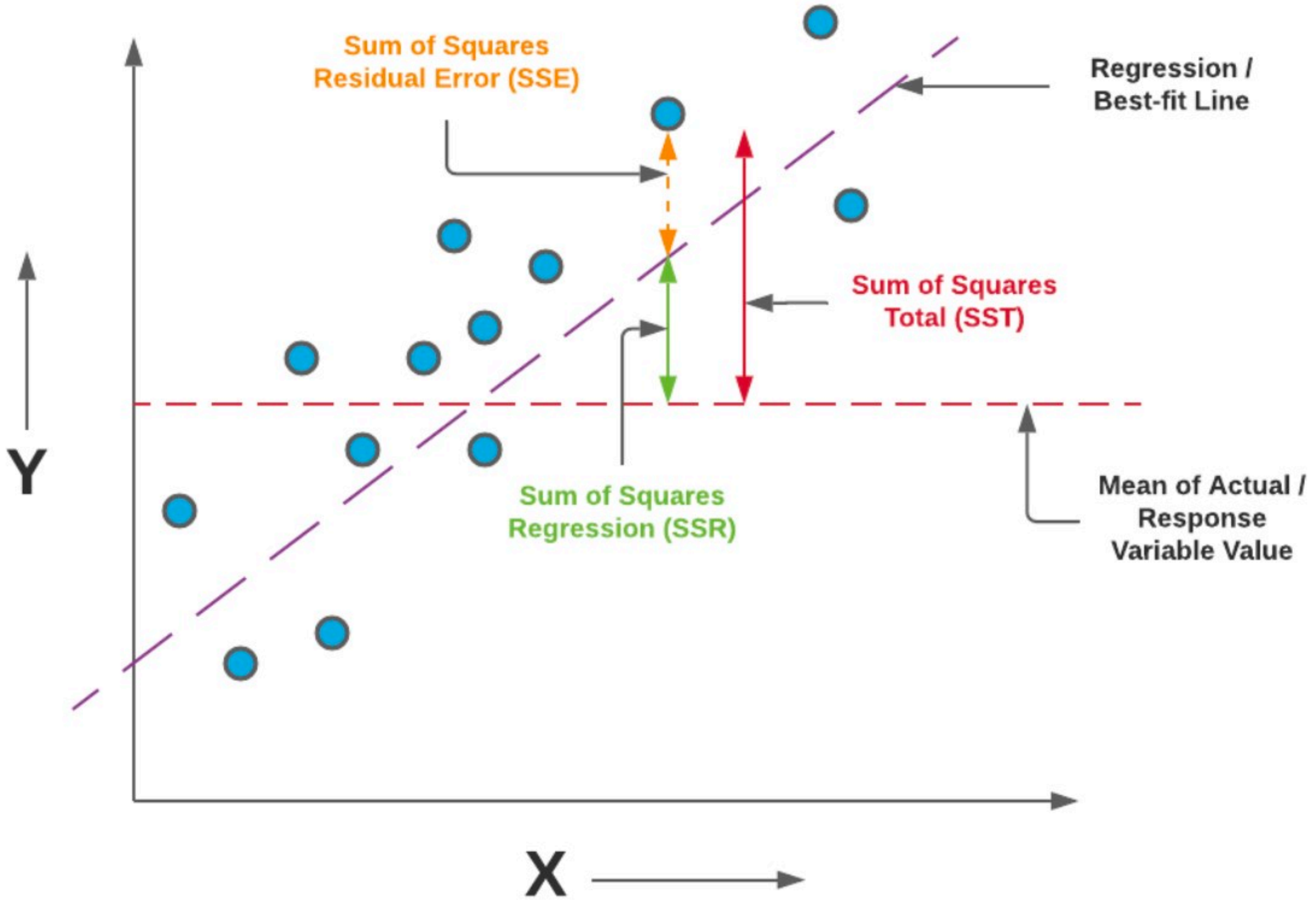
Least Squares Solution

Find the least squares solution to $\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \vec{x} = \begin{bmatrix} 1 \\ 3 \\ 8 \\ 2 \end{bmatrix}$.

R2 Score

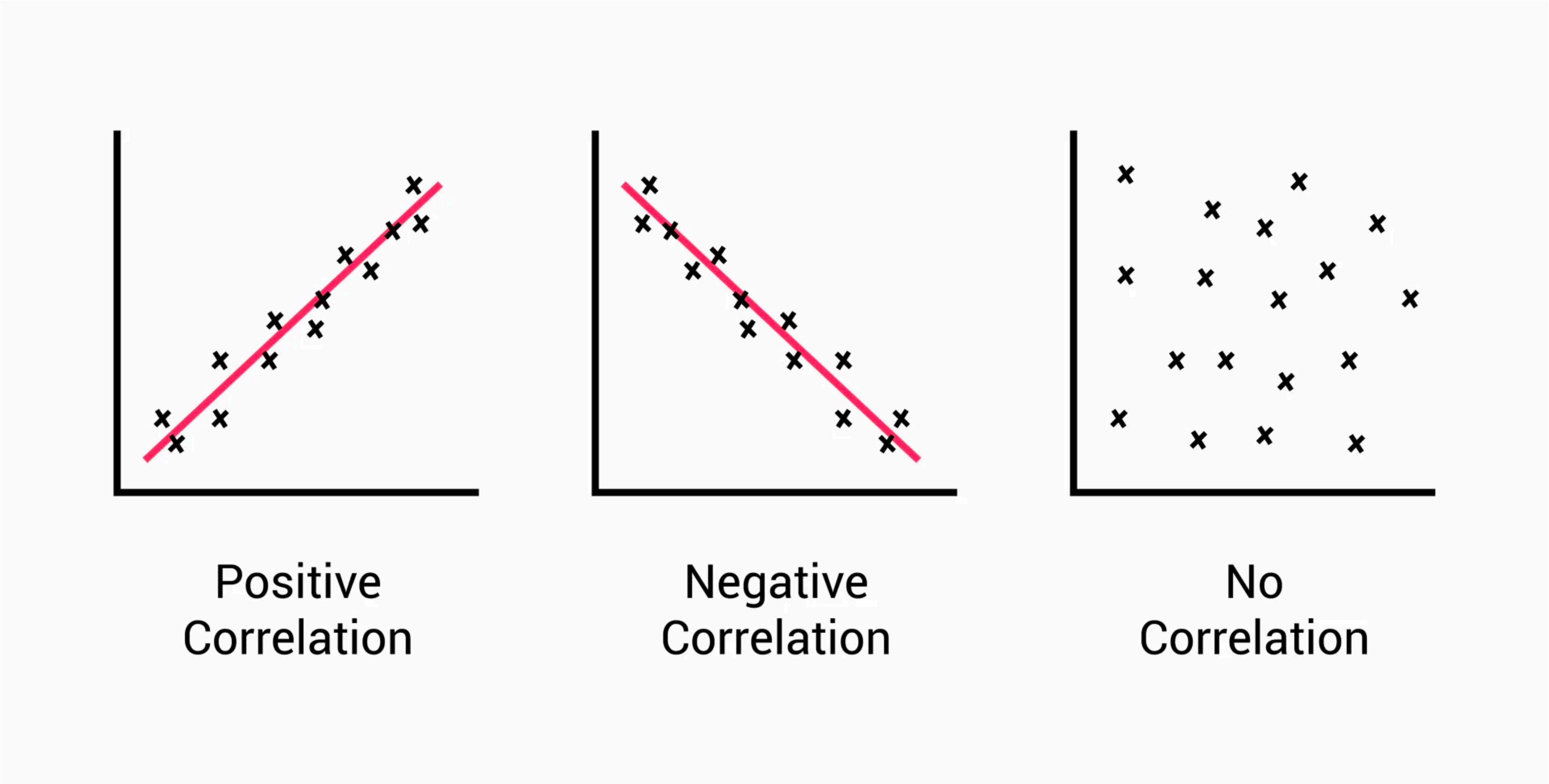
$$R2\ score = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad R2\ score = 1 - \frac{SS_R}{SS_M}$$





Sum of Squares Residual Error (SSE)
Regression / Best-fit Line
Sum of Squares Total (SST)
Y
Sum of Squares Regression (SSR)
Mean of Actual / Response Variable Value
X

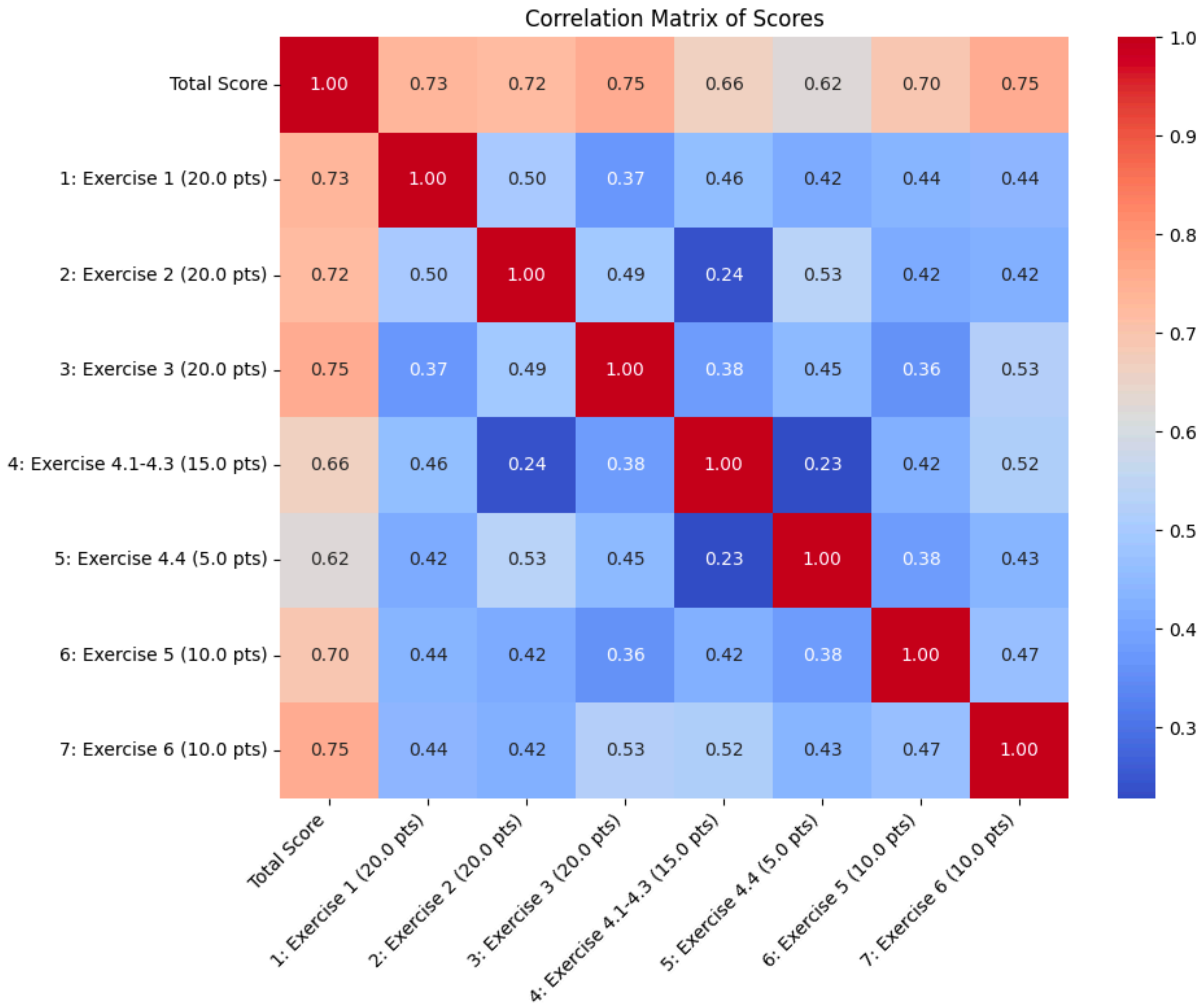
Correlation

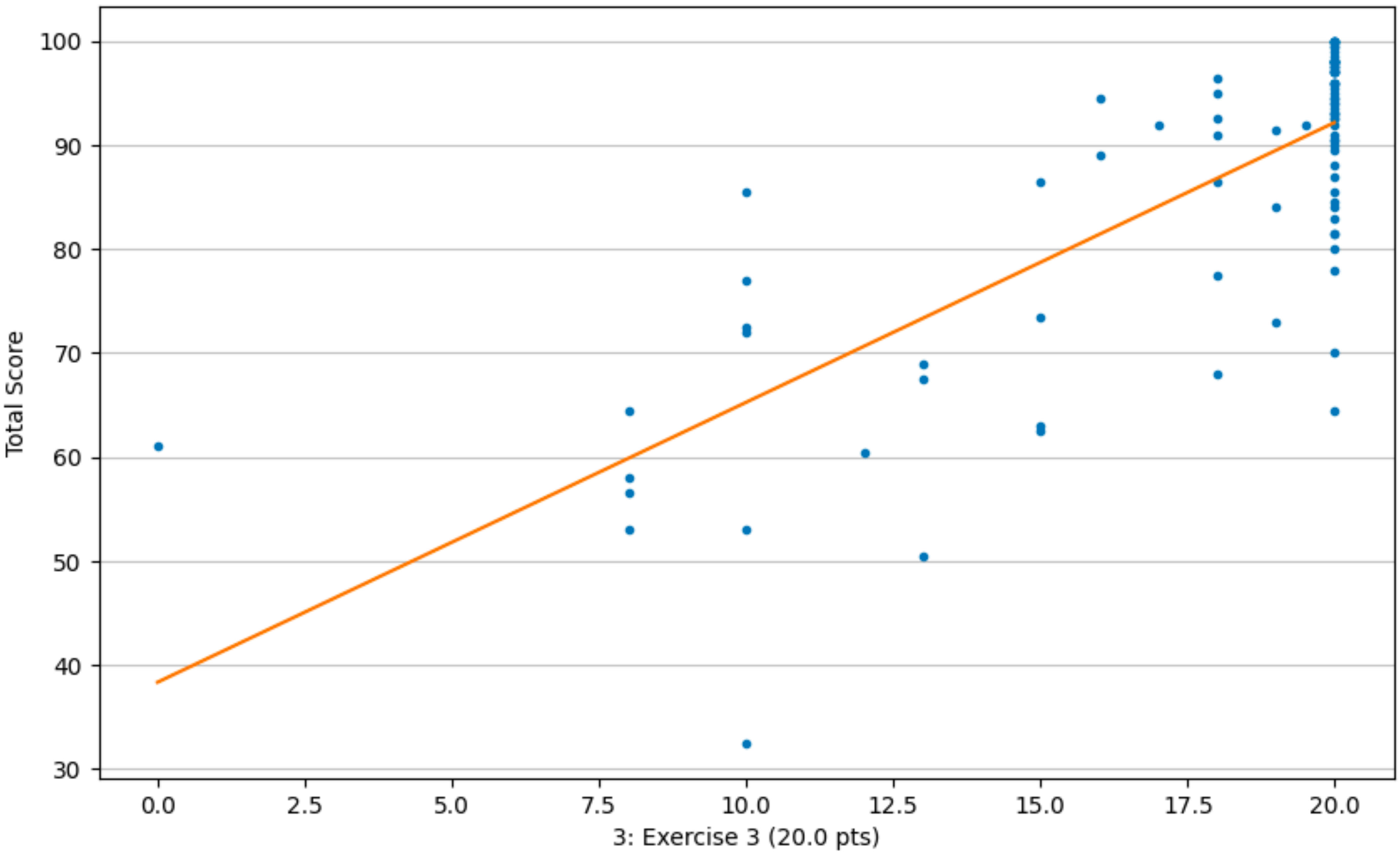


Correlation = R2 Square for only 1 feature

- **Thm** Correlation=R2 square when we only 1 feature to do linear regression

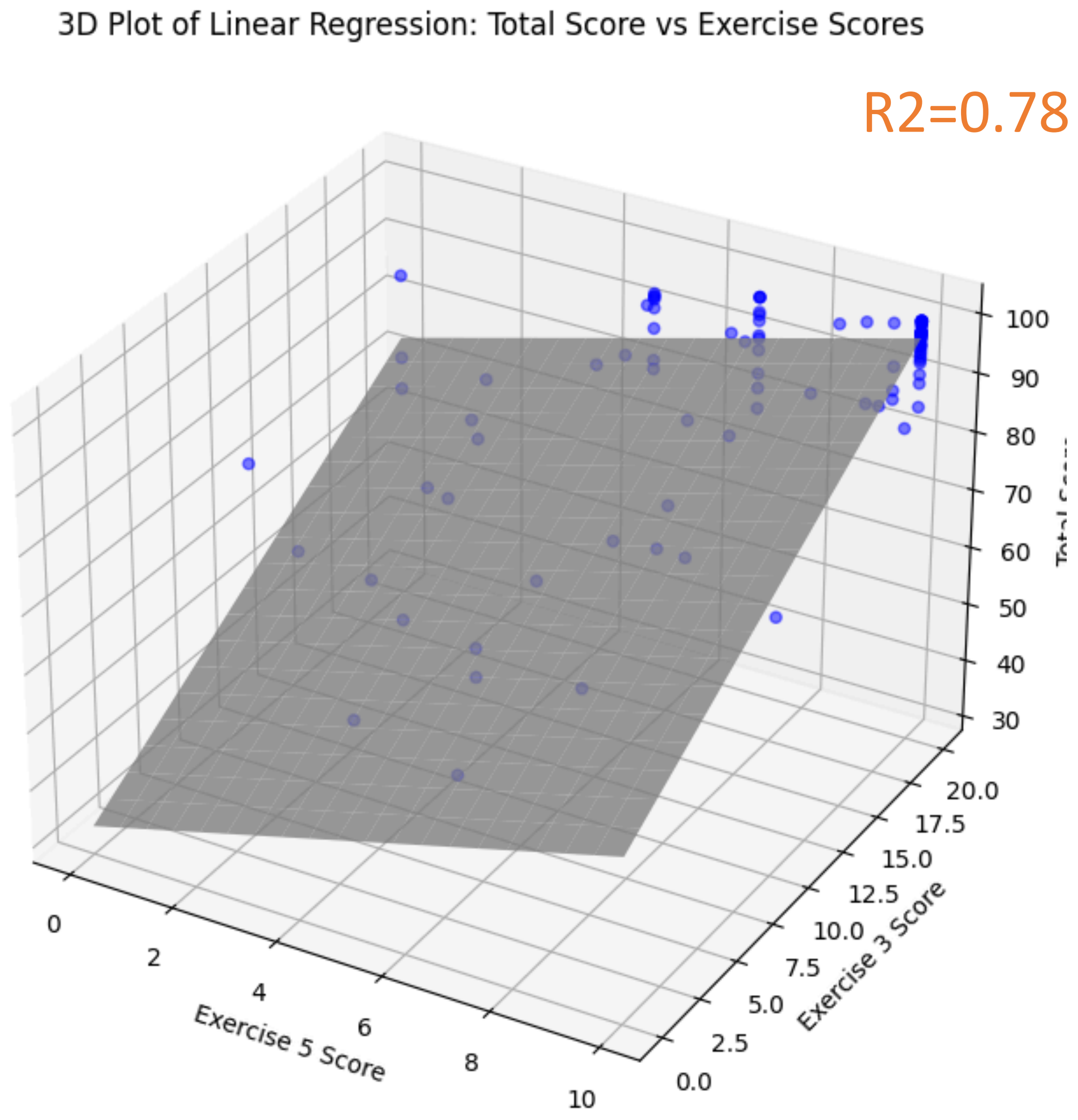
You midterm score

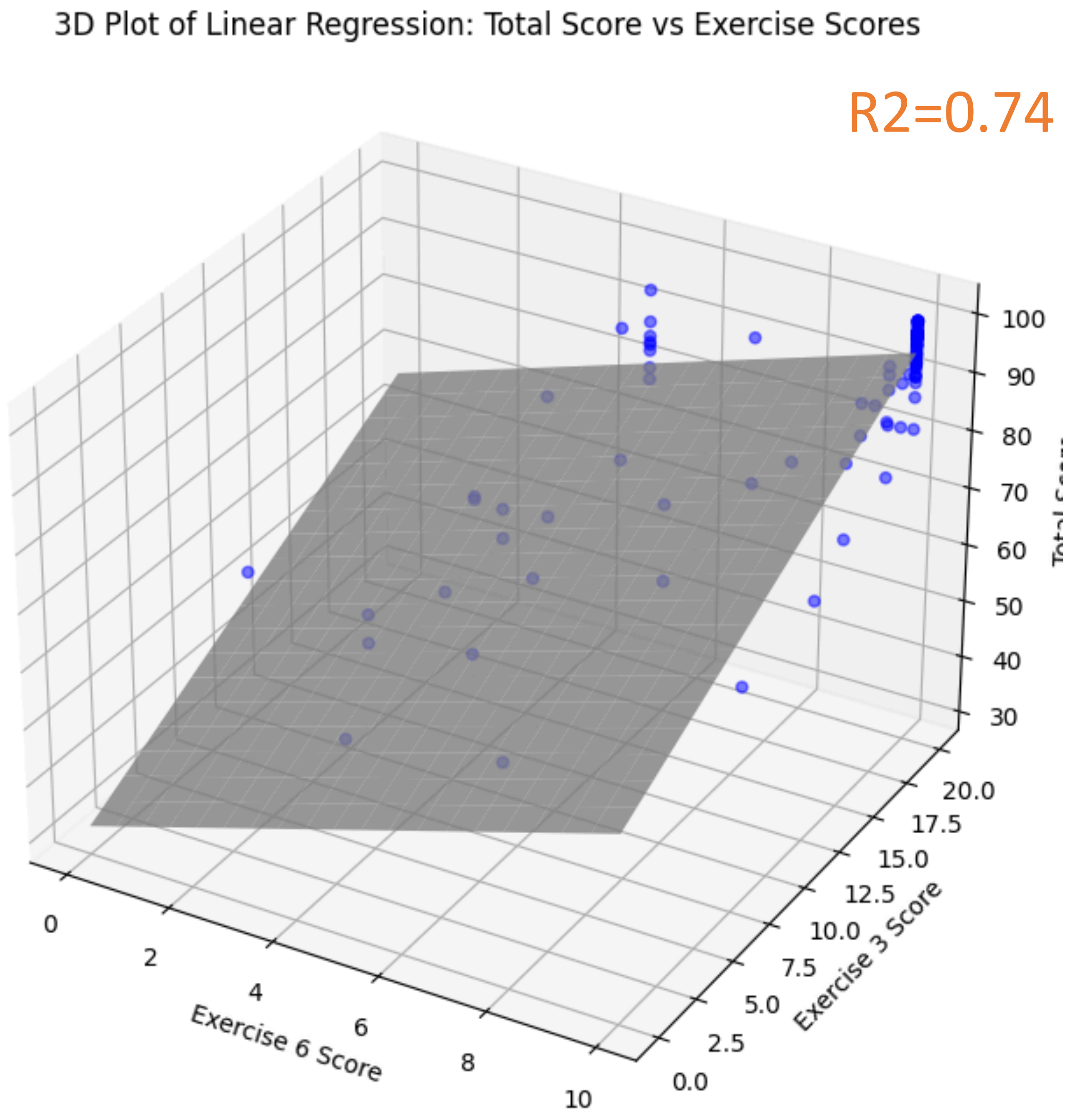




100
90
80
70
60
50
40
30
Total Score
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
3: Exercise 3 (20.0 pts)

Exercise 5 and 6 which is *more powerful*?





Try to Design a Linear Algebra Test using Linear Algebra

```python
# Preparing the data
feature_columns = ['1: Exercise 1 (20.0 pts)', '2: Exercise 2 (20.0 pts)',
                   '3: Exercise 3 (20.0 pts)', '4: Exercise 4.1-4.3 (15.0 pts)',
                   '5: Exercise 4.4 (5.0 pts)', '6: Exercise 5 (10.0 pts)',
                   '7: Exercise 6 (10.0 pts)']
X_features = scores_data[feature_columns]
y_total_score = scores_data['Total Score']

# Standardizing the features
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X_features)

# Splitting the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y_total_score, test_size=0.2, random_state=42)

# Fitting the Lasso regression model
lasso = Lasso(alpha=5)
lasso.fit(X_train, y_train)

# Getting the coefficients
lasso_coefficients = lasso.coef_

# Displaying the coefficients
coefficients_df = pd.DataFrame({'Feature': feature_columns, 'Coefficient': lasso_coefficients})
coefficients_df
```

	Feature	Coefficient
0	1: Exercise 1 (20.0 pts)	0.324832
1	2: Exercise 2 (20.0 pts)	0.000000
2	3: Exercise 3 (20.0 pts)	3.082411
3	4: Exercise 4.1-4.3 (15.0 pts)	1.556173
4	5: Exercise 4.4 (5.0 pts)	0.000000
5	6: Exercise 5 (10.0 pts)	1.712245
6	7: Exercise 6 (10.0 pts)	2.447044

Lasso:
best linear fit with possible fewer entries

Larger alpha leads to more zeros!
(Means less problems in exam can know
Your’s status of learning!)

Try to Design a Linear Algebra Test using Linear Algebra

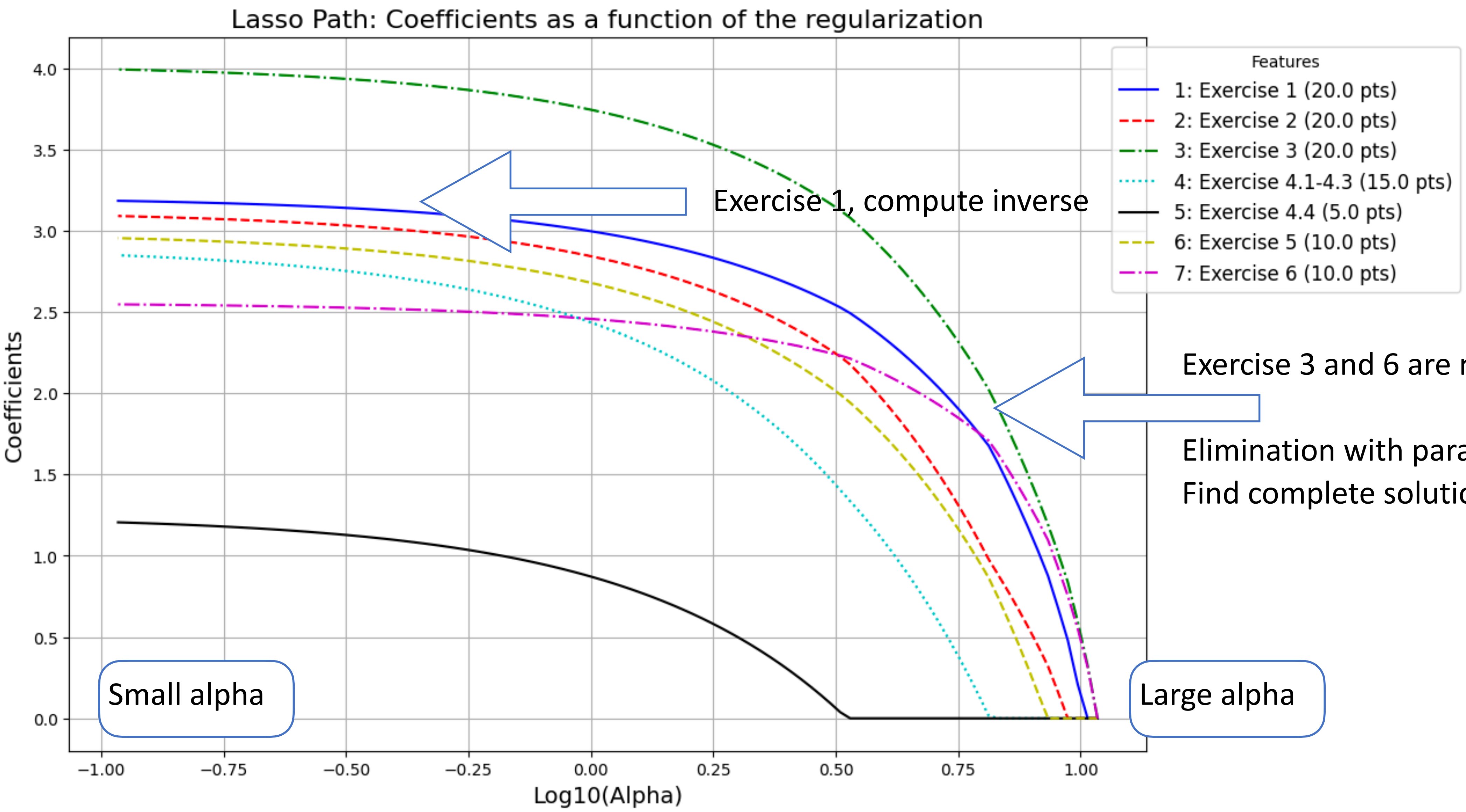
```python
# Preparing the data
feature_columns = ['1: Exercise 1 (20.0 pts)', '2: Exercise 2 (20.0 pts)',
                   '3: Exercise 3 (20.0 pts)', '4: Exercise 4.1-4.3 (15.0 pts)',
                   '5: Exercise 4.4 (5.0 pts)', '6: Exercise 5 (10.0 pts)',
                   '7: Exercise 6 (10.0 pts)']
X_features = scores_data[feature_columns]
y_total_score = scores_data['Total Score']

# Standardizing the features
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X_features)

# Splitting the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y_total_score, test_size=0.2, random_state=42)

# Fitting the Lasso regression model
lasso = Lasso(alpha=10)
lasso.fit(X_train, y_train)

# Getting the coefficients
lasso_coefficients = lasso.coef_

# Displaying the coefficients
coefficients_df = pd.DataFrame({'Feature': feature_columns, 'Coefficient': lasso_coefficients})
coefficients_df
```

	Feature	Coefficient
0	1: Exercise 1 (20.0 pts)	0.000000
1	2: Exercise 2 (20.0 pts)	0.000000
2	3: Exercise 3 (20.0 pts)	0.698581
3	4: Exercise 4.1-4.3 (15.0 pts)	0.000000
4	5: Exercise 4.4 (5.0 pts)	0.000000
5	6: Exercise 5 (10.0 pts)	0.000000
6	7: Exercise 6 (10.0 pts)	0.883778

Lasso:
best linear fit with possible fewer entries

Larger alpha leads to more zeros!
(Means less problems in exam can know
Your's status of learning!)

Try to Design a Linear Algebra Test using Linear Algebra



Exercise 3 and 6 are most important

Elimination with parameterized matrix

Find complete solution of Ax=b

Worked Example – Best Fit Ellipse

Best Fit Ellipse

Find the best fit ellipse for the points $(0, 2)$, $(2, 1)$, $(1, -1)$, $(-1, -2)$, $(-3, 1)$.

The general equation for an ellipse is

$$x^2 + Ay^2 + Bxy + Cx + Dy + E = 0$$

Best Fit Ellipse

Find the best fit ellipse for the points $(0,2)$, $(2,1)$, $(1,-1)$, $(-1,-2)$, $(-3,1)$.

The general equation for an ellipse is

$$x^2 + Ay^2 + Bxy + Cx + Dy + E = 0$$

So we want to solve:

$$\begin{aligned}
(0)^2 + \quad A(2)^2 + \quad B(0)(2) + \quad C(0) + \quad D(2) + E &= 0 \\
(2)^2 + \quad A(1)^2 + \quad B(2)(1) + \quad C(2) + \quad D(1) + E &= 0 \\
(1)^2 + A(-1)^2 + \quad B(1)(-1) + \quad C(1) + D(-1) + E &= 0 \\
(-1)^2 + A(-2)^2 + B(-1)(-2) + C(-1) + D(-2) + E &= 0 \\
(-3)^2 + \quad A(1)^2 + \quad B(-3)(1) + C(-3) + \quad D(1) + E &= 0
\end{aligned}$$

Best Fit Ellipse

Find the best fit ellipse for the points $(0,2)$, $(2,1)$, $(1,-1)$, $(-1,-2)$, $(-3,1)$.

The general equation for an ellipse is

$$x^2 + Ay^2 + Bxy + Cx + Dy + E = 0$$

So we want to solve:

$$\begin{aligned}
(0)^2 + A(2)^2 + B(0)(2) + C(0) + D(2) + E &= 0\\
(2)^2 + A(1)^2 + B(2)(1) + C(2) + D(1) + E &= 0\\
(1)^2 + A(-1)^2 + B(1)(-1) + C(1) + D(-1) + E &= 0\\
(-1)^2 + A(-2)^2 + B(-1)(-2) + C(-1) + D(-2) + E &= 0\\
(-3)^2 + A(1)^2 + B(-3)(1) + C(-3) + D(1) + E &= 0
\end{aligned}$$

In matrix form:

$$\begin{pmatrix} 4 & 0 & 0 & 2 & 1\\ 1 & 2 & 2 & 1 & 1\\ 1 & -1 & 1 & -1 & 1\\ 4 & 2 & -1 & -2 & 1\\ 1 & -3 & -3 & 1 & 1 \end{pmatrix} \begin{pmatrix} A\\ B\\ C\\ D\\ E \end{pmatrix} = \begin{pmatrix} 0\\ -4\\ -1\\ -1\\ -9 \end{pmatrix}.$$

Best Fit Ellipse

$$A = \begin{pmatrix} 4 & 0 & 0 & 2 & 1 \\ 1 & 2 & 2 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 4 & 2 & -1 & -2 & 1 \\ 1 & -3 & -3 & 1 & 1 \end{pmatrix} \qquad b = \begin{pmatrix} 0 \\ -4 \\ -1 \\ -1 \\ -9 \end{pmatrix}.$$

$$A^T A = \begin{pmatrix} 35 & 6 & -4 & 1 & 11 \\ 6 & 18 & 10 & -4 & 0 \\ -4 & 10 & 15 & 0 & -1 \\ 1 & -4 & 0 & 11 & 1 \\ 11 & 0 & -1 & 1 & 5 \end{pmatrix} \qquad A^T b = \begin{pmatrix} -18 \\ 18 \\ 19 \\ -10 \\ -15 \end{pmatrix}$$

Best Fit Ellipse

$$A = \begin{pmatrix} 4 & 0 & 0 & 2 & 1 \\ 1 & 2 & 2 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 4 & 2 & -1 & -2 & 1 \\ 1 & -3 & -3 & 1 & 1 \end{pmatrix} \qquad b = \begin{pmatrix} 0 \\ -4 \\ -1 \\ -1 \\ -9 \end{pmatrix}.$$

$$A^T A = \begin{pmatrix} 35 & 6 & -4 & 1 & 11 \\ 6 & 18 & 10 & -4 & 0 \\ -4 & 10 & 15 & 0 & -1 \\ 1 & -4 & 0 & 11 & 1 \\ 11 & 0 & -1 & 1 & 5 \end{pmatrix} \qquad A^T b = \begin{pmatrix} -18 \\ 18 \\ 19 \\ -10 \\ -15 \end{pmatrix}$$

Row reduce:

$$\left(\begin{array}{ccccc|c} 35 & 6 & -4 & 1 & 11 & -18 \\ 6 & 18 & 10 & -4 & 0 & 18 \\ -4 & 10 & 15 & 0 & -1 & 19 \\ 1 & -4 & 0 & 11 & 1 & -10 \\ 11 & 0 & -1 & 1 & 5 & -15 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & 0 & 16/7 \\ 0 & 1 & 0 & 0 & 0 & -8/7 \\ 0 & 0 & 1 & 0 & 0 & 15/7 \\ 0 & 0 & 0 & 1 & 0 & -6/7 \\ 0 & 0 & 0 & 0 & 1 & -52/7 \end{array}\right)$$

Best Fit Ellipse

$$A = \begin{pmatrix} 4 & 0 & 0 & 2 & 1 \\ 1 & 2 & 2 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 4 & 2 & -1 & -2 & 1 \\ 1 & -3 & -3 & 1 & 1 \end{pmatrix} \qquad b = \begin{pmatrix} 0 \\ -4 \\ -1 \\ -1 \\ -9 \end{pmatrix}.$$

$$A^T A = \begin{pmatrix} 35 & 6 & -4 & 1 & 11 \\ 6 & 18 & 10 & -4 & 0 \\ -4 & 10 & 15 & 0 & -1 \\ 1 & -4 & 0 & 11 & 1 \\ 11 & 0 & -1 & 1 & 5 \end{pmatrix} \qquad A^T b = \begin{pmatrix} -18 \\ 18 \\ 19 \\ -10 \\ -15 \end{pmatrix}$$

Row reduce:

$$\left(\begin{array}{ccccc|c} 35 & 6 & -4 & 1 & 11 & -18 \\ 6 & 18 & 10 & -4 & 0 & 18 \\ -4 & 10 & 15 & 0 & -1 & 19 \\ 1 & -4 & 0 & 11 & 1 & -10 \\ 11 & 0 & -1 & 1 & 5 & -15 \end{array}\right) \rightsquigarrow \left(\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & 0 & 16/7 \\ 0 & 1 & 0 & 0 & 0 & -8/7 \\ 0 & 0 & 1 & 0 & 0 & 15/7 \\ 0 & 0 & 0 & 1 & 0 & -6/7 \\ 0 & 0 & 0 & 0 & 1 & -52/7 \end{array}\right)$$

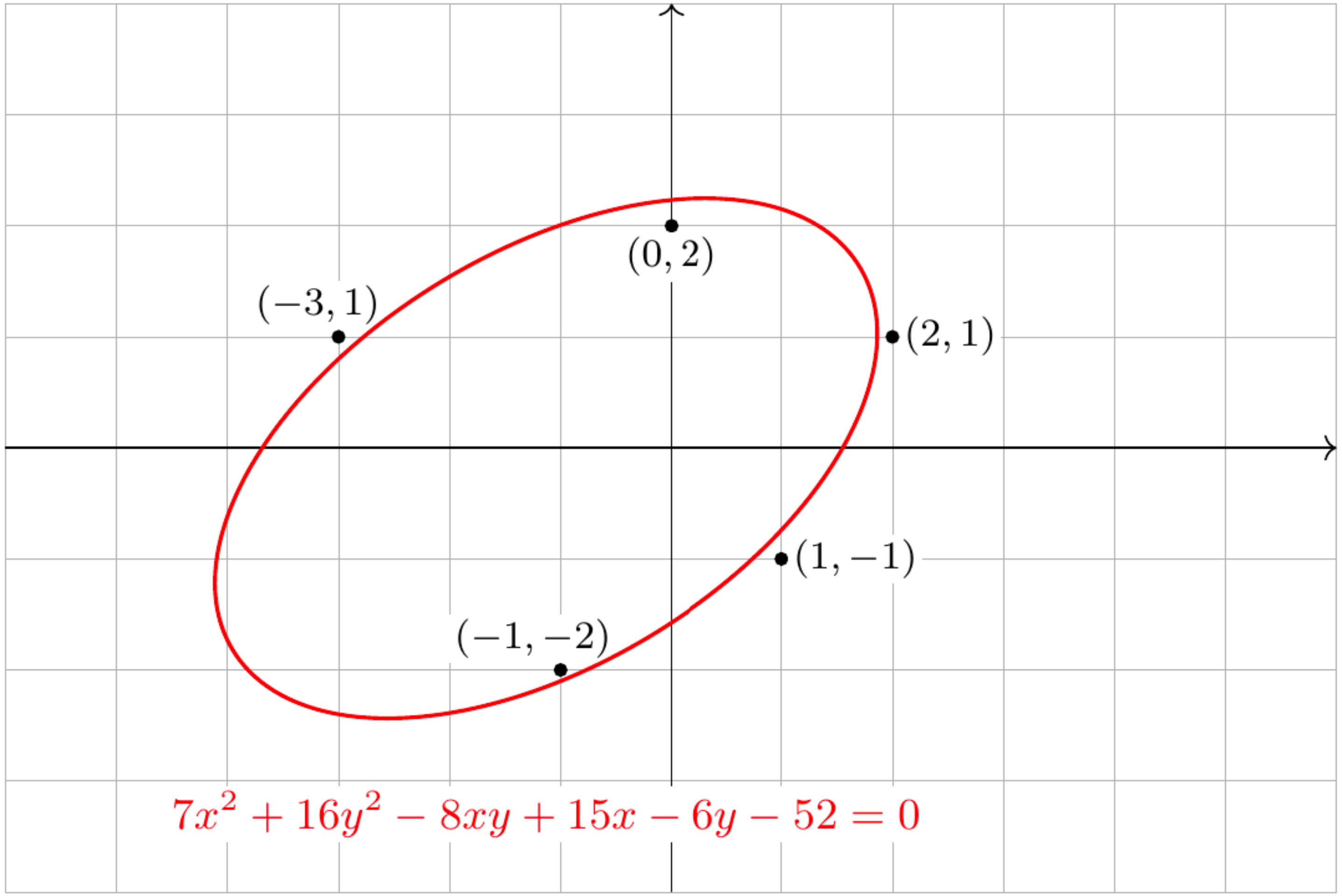
Best fit ellipse:

$$x^2 + \frac{16}{7}y^2 - \frac{8}{7}xy + \frac{15}{7}x - \frac{6}{7}y - \frac{52}{7} = 0$$

or

$$7x^2 + 16y^2 - 8xy + 15x - 6y - 52 = 0.$$

Best Fit Ellipse



Remark: Gauss invented the method of least squares to do exactly this: he predicted the (elliptical) orbit of the asteroid Ceres as it passed behind the sun in 1801.

Exercises

If you want more suggestions from the book (solutions easily available), message on the corresponding Campuswire thread

Exercises

Find the best-fit line $b = C + Dt$ through the points $(1, 1)$, $(2, 5)$, and $(-1, 2)$.

Exercises

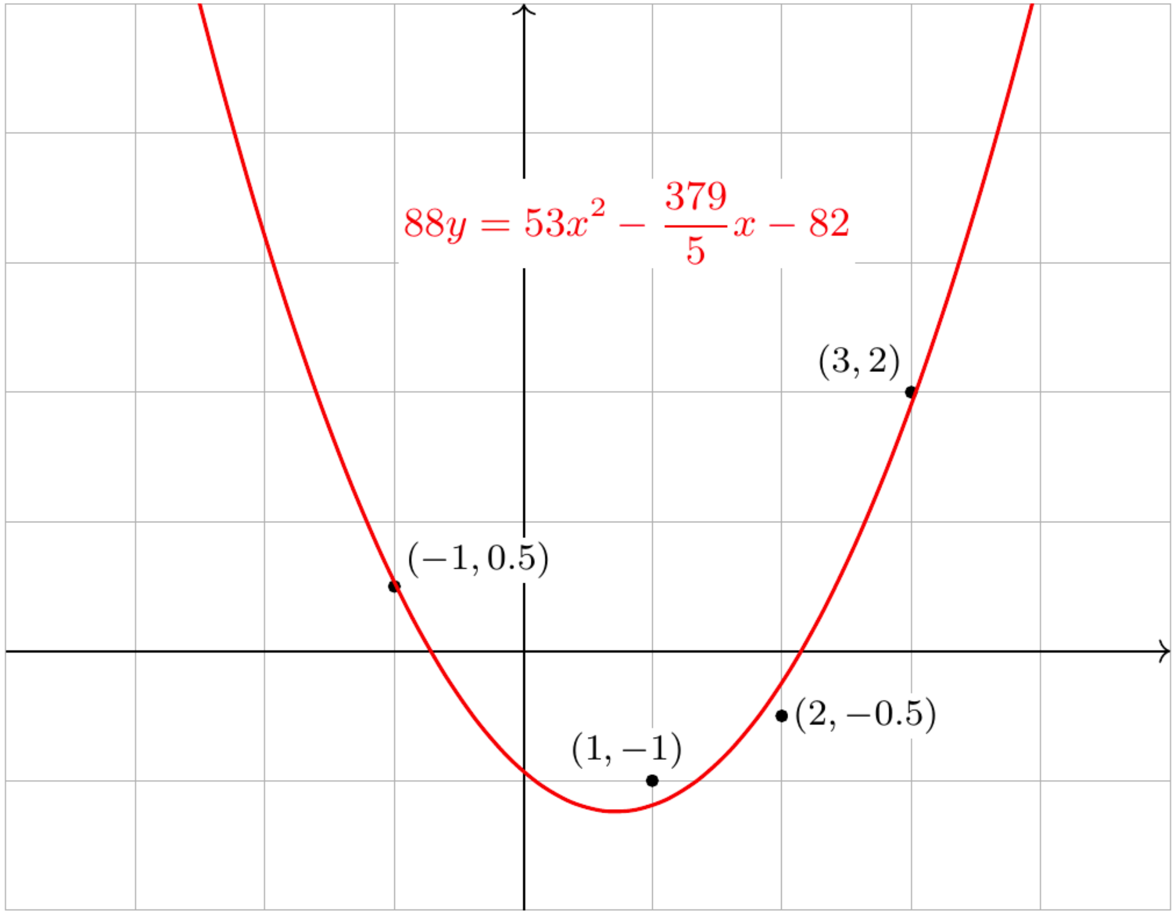
Find the projection of $(2, 3, -2, 1)$ onto the nullspace of $A = \begin{bmatrix} 1 & 3 & 1 & 2 \\ 2 & 6 & 4 & 8 \\ 0 & 0 & 2 & 4 \end{bmatrix}$.

Exercises

What least squares problem $Ax = b$ finds the best parabola through the points $(-1, 0.5)$, $(1, -1)$, $(2, -0.5)$, $(3, 2)$?

Exercises

What least squares problem $Ax = b$ finds the best parabola through the points $(-1, 0.5)$, $(1, -1)$, $(2, -0.5)$, $(3, 2)$?



Questions?