



11.9 Linear regression and least squares





Introduction

- In this topic, we will
 - Find the best approximation of a target vector as a linear combination of a given set of vectors
 - Deduce the normal equations that must be solved to find these coefficients
 - Look at two examples





Problem

- Suppose that we would like to find if a linear combination of n linearly independent m -dimensional vectors that equals a given target vector $\mathbf{y}$
 - For example, suppose we have $\mathbf{v}_1, \dots, \mathbf{v}_n$ where $n \leq m$
- We can define $V = (\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3 \ \cdots \ \mathbf{v}_n)$ and try to solve:
$$\begin{pmatrix} V & \vdots & \mathbf{y} \end{pmatrix}$$
- Because the vectors are linearly independent, if a solution exists, it will be unique





Previous solutions

- What if no solution exists?
 - Could we find a best approximation?
 - That is, what linear combination of the vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ best approximate the vector $\mathbf{y}$?
- Recall that if we have a single vector $\mathbf{v}$,
then the best approximation of $\mathbf{y}$ is $\text{proj}_{\mathbf{v}}(\mathbf{y}) = \frac{\langle \mathbf{v}, \mathbf{y} \rangle}{\langle \mathbf{v}, \mathbf{v} \rangle} \mathbf{v}$
 - That is, the projection of $\mathbf{y}$ onto $\mathbf{v}$
- Recall also that if we had an orthonormal collection of vectors $\hat{\mathbf{v}}_1, \dots, \hat{\mathbf{v}}_n$, the the best approximation of $\mathbf{y}$ is
$$\langle \hat{\mathbf{v}}_1, \mathbf{y} \rangle \hat{\mathbf{v}}_1 + \langle \hat{\mathbf{v}}_2, \mathbf{y} \rangle \hat{\mathbf{v}}_2 + \dots + \langle \hat{\mathbf{v}}_n, \mathbf{y} \rangle \hat{\mathbf{v}}_n$$





Finding a solution

- Apart from these two special cases, what else can we do?
 - Idea: recall that $\text{proj}_{\mathbf{v}}(\mathbf{y})$ is that scalar multiple of $\mathbf{v}$ that is closest to the vector $\mathbf{y}$
 - Note also that the error $\mathbf{y} - \text{proj}_{\mathbf{v}}(\mathbf{y})$ is perpendicular to $\mathbf{v}$
- Now, the range of $V = (\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3 \ \cdots \ \mathbf{v}_n)$ is all linear combinations of the given vectors
 - This forms a subspace in $\mathbf{R}^m$
 - Can we find the one vector in $\text{range}(V)$ that is closest to $\mathbf{y}$?

- That is, find a solution vector $\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}$ where $V\boldsymbol{\alpha}$ is closest to $\mathbf{y}$





Finding a solution

- It is actually rather difficult to solve this in general

$$\|V\alpha - \mathbf{y}\|_2$$

even if you minimize $\|V\alpha - \mathbf{y}\|_2^2 = \langle V\alpha - \mathbf{y}, V\alpha - \mathbf{y} \rangle$

- Instead, consider the following approach:
 - Find that α such that $V\alpha - \mathbf{y}$ that is perpendicular to $\text{range}(V)$
 - This might sound familiar...





Finding a solution

- Now, $V: \mathbf{F}^n \rightarrow \mathbf{F}^m$, so for $V\mathbf{a} - \mathbf{y}$ to be perpendicular to every point in the range of V , we require that

$$\langle V\mathbf{u}, V\mathbf{a} - \mathbf{y} \rangle = 0$$

for all vectors $\mathbf{u} \in \mathbf{F}^n$

- Recall that we can bring the linear operator V to the right-hand side by using the adjoint, so

$$\langle \mathbf{u}, V^*(V\mathbf{a} - \mathbf{y}) \rangle = 0$$

- This says that $V^*(V\mathbf{a} - \mathbf{y})$ is perpendicular to $\mathbf{u}$
for every vector $\mathbf{u} \in \mathbf{F}^n$
 - There is only one vector in $\mathbf{F}^n$ that is perpendicular to all vectors in that vector space, and that is the zero vector





Finding a solution

- Thus, if $V: \mathbf{F}^n \rightarrow \mathbf{F}^m$ and $\langle \mathbf{u}, V^* (V\mathbf{a} - \mathbf{y}) \rangle = 0$ for all $\mathbf{u} \in \mathbf{F}^n$, the only possibility is that

$$V^* (V\mathbf{a} - \mathbf{y}) = \mathbf{0}_n$$

- Expanding this, we have

$$V^* V\mathbf{a} - V^* \mathbf{y} = \mathbf{0}_n$$

$$V^* V\mathbf{a} = V^* \mathbf{y}$$

- Because the vectors in V were linearly independent, it follows that $\text{rank}(V^* V) = \text{rank}(V) = n$, so a solution is guaranteed to exist
- This solution is called the *least squares* solutions because it minimizes $\|V\mathbf{a} - \mathbf{y}\|_2$





Example

- For example, given

$$\mathbf{v}_1 = \begin{pmatrix} -1 \\ 3 \\ 1 \\ -3 \end{pmatrix} \text{ and } \mathbf{v}_2 = \begin{pmatrix} -3 \\ 1 \\ 1 \\ -1 \end{pmatrix}$$

- What linear combination of these two vectors best approximates this vector?

$$\mathbf{y} = \begin{pmatrix} -8 \\ 2 \\ -2 \\ 6 \end{pmatrix} \quad \begin{pmatrix} -1 & -3 & \vdots & -8 \\ 3 & 1 & \vdots & 2 \\ 1 & 1 & \vdots & -2 \\ -3 & -1 & \vdots & 6 \end{pmatrix} \sim \begin{pmatrix} -1 & -3 & \vdots & -8 \\ 0 & -8 & \vdots & -22 \\ 0 & 0 & \vdots & -4.5 \\ 0 & 0 & \vdots & 0 \end{pmatrix}$$





Example

- Now, if $V = \begin{pmatrix} -1 & -3 \\ 3 & 1 \\ 1 & 1 \\ -3 & -1 \end{pmatrix}$, we must solve $V^T V \boldsymbol{\alpha} = V^T \mathbf{y}$ where $\mathbf{y} = \begin{pmatrix} -8 \\ 2 \\ -2 \\ 6 \end{pmatrix}$

$$V^T V = \begin{pmatrix} -1 & 3 & 1 & -3 \\ -3 & 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} -1 & -3 \\ 3 & 1 \\ 1 & 1 \\ -3 & -1 \end{pmatrix} = \begin{pmatrix} 20 & 10 \\ 10 & 12 \end{pmatrix}$$

$$V^T \mathbf{y} = \begin{pmatrix} -1 & 3 & 1 & -3 \\ -3 & 1 & 1 & -1 \end{pmatrix} \begin{pmatrix} -8 \\ 2 \\ -2 \\ 6 \end{pmatrix} = \begin{pmatrix} -6 \\ 18 \end{pmatrix}$$

$$\left(\begin{array}{cc|c} 20 & 10 & -6 \\ 10 & 12 & 18 \end{array} \right) \sim \left(\begin{array}{cc|c} 20 & 10 & -6 \\ 0 & 7 & 21 \end{array} \right)$$

$$\boldsymbol{\alpha} = \begin{pmatrix} -1.8 \\ 3 \end{pmatrix}$$





Example

- Thus, the best approximation of $\mathbf{y} = \begin{pmatrix} -8 \\ 2 \\ -2 \\ 6 \end{pmatrix}$ using $\begin{pmatrix} -1 \\ 3 \\ 1 \\ -3 \end{pmatrix}$, $\begin{pmatrix} -3 \\ 1 \\ 1 \\ -1 \end{pmatrix}$

$$\text{is } -1.8 \begin{pmatrix} -1 \\ 3 \\ 1 \\ -3 \end{pmatrix} + 3 \begin{pmatrix} -3 \\ 1 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -7.2 \\ -2.4 \\ 1.2 \\ 2.4 \end{pmatrix}$$

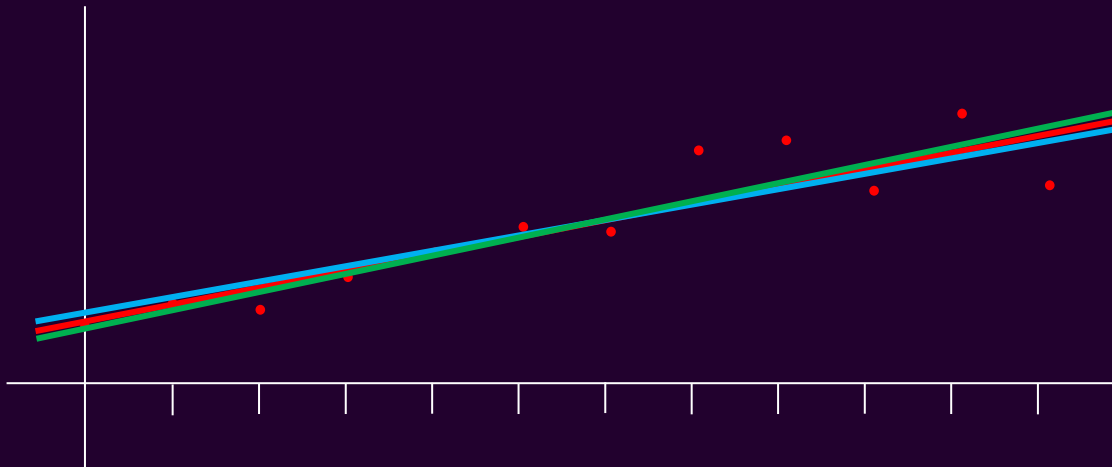
- This isn't "close", but any other linear combination will be further away from the target vector





Applications

- Suppose you are periodically reading a sensor and it is returning a scalar of some significance
 - Question: what line $at + b$ best fits this data for then you know the best estimate of the current rate of change: a

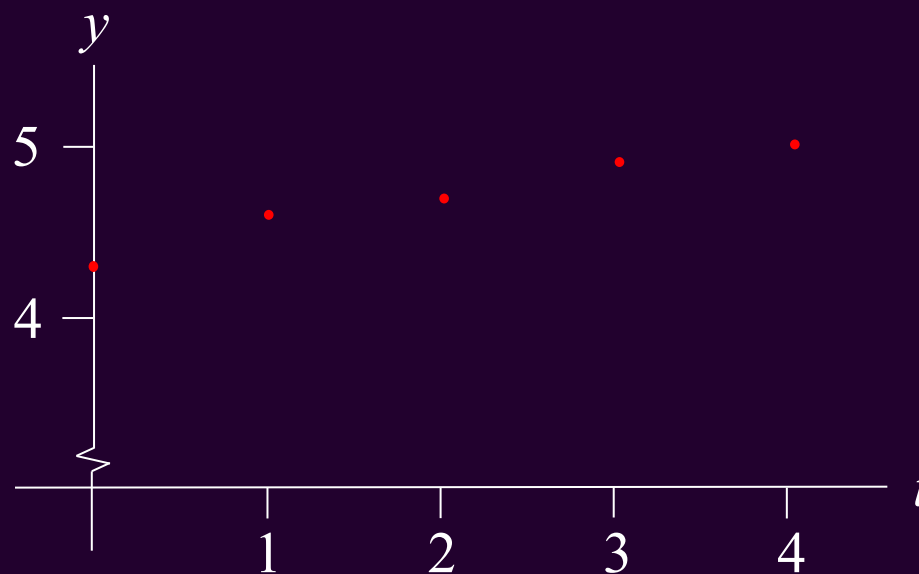




Applications

- Consider this data

$(0, 4.3), (1, 4.6), (2, 4.7), (3, 4.9), (4, 5.0)$





Example

- Thus, we have these pairs of t - y points:

$$(0, 4.3), (1, 4.6), (2, 4.7), (3, 4.9), (4, 5.0)$$

- Suppose we want to find the best fitting line $y = at + b$
 - Thus, we have five equations in two unknowns:

$$a \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} + b \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \approx \begin{pmatrix} 4.3 \\ 4.6 \\ 4.7 \\ 4.9 \\ 5.0 \end{pmatrix}$$

$$0a + b = 4.3$$

$$1a + b = 4.6$$

$$2a + b = 4.7$$

$$3a + b = 4.9$$

$$4a + b = 5.0$$

$$\mathbf{y} = \begin{pmatrix} 4.3 \\ 4.6 \\ 4.7 \\ 4.9 \\ 5.0 \end{pmatrix}$$





Example

- Thus, we have these pairs of t - y points:

$(0, 4.3), (1, 4.6), (2, 4.7), (3, 4.9), (4, 5.0)$

- Therefore, we must define

$$V = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix}$$

$$0a + b = 4.3$$

$$1a + b = 4.6$$

$$2a + b = 4.7$$

$$3a + b = 4.9$$

$$4a + b = 5.0$$

$$\mathbf{y} = \begin{pmatrix} 4.3 \\ 4.6 \\ 4.7 \\ 4.9 \\ 5.0 \end{pmatrix}$$





Example

- Next we calculate:

$$V^T V = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \end{pmatrix} = \begin{pmatrix} 30 & 10 \\ 10 & 5 \end{pmatrix}$$
$$V^T \mathbf{y} = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 4.3 \\ 4.6 \\ 4.7 \\ 4.9 \\ 5.0 \end{pmatrix} = \begin{pmatrix} 48.7 \\ 23.5 \end{pmatrix}$$
$$\begin{pmatrix} 30 & 10 \\ 10 & 5 \end{pmatrix} \boldsymbol{\alpha} = \begin{pmatrix} 48.7 \\ 23.5 \end{pmatrix}$$
$$\boldsymbol{\alpha} = \begin{pmatrix} 0.17 \\ 4.36 \end{pmatrix}$$

- Thus, the best-fitting line is $y = 0.17t + 4.36$

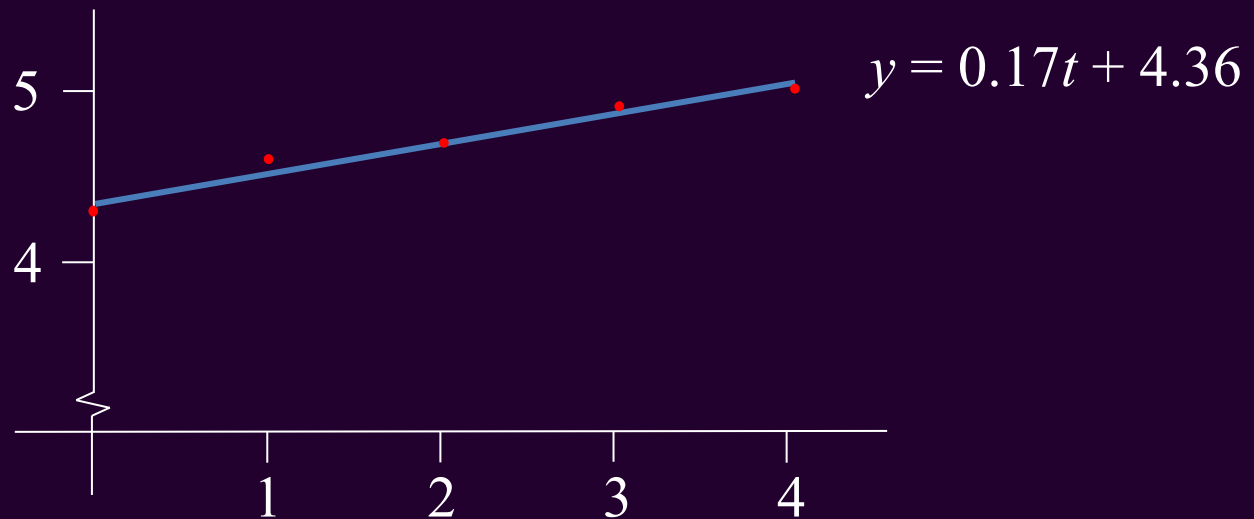




Applications

- Thus, looking at this result:

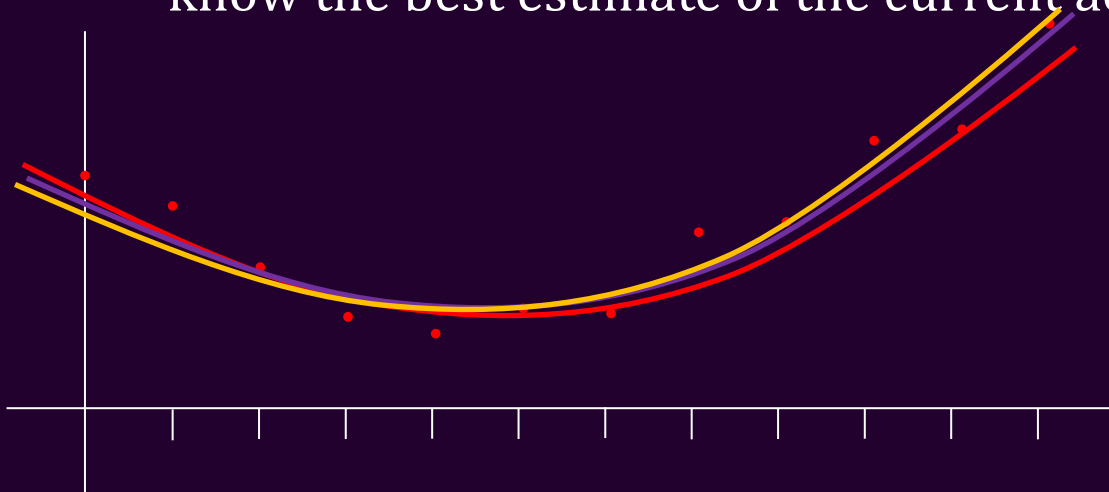
$(0, 4.3), (1, 4.6), (2, 4.7), (3, 4.9), (4, 5.0)$





Applications

- Suppose you are periodically reading a sensor and it is returning a scalar of some significance
 - Question: what curve $at^2 + bt + c$ best fits this data for then you know the best estimate of the current acceleration: a

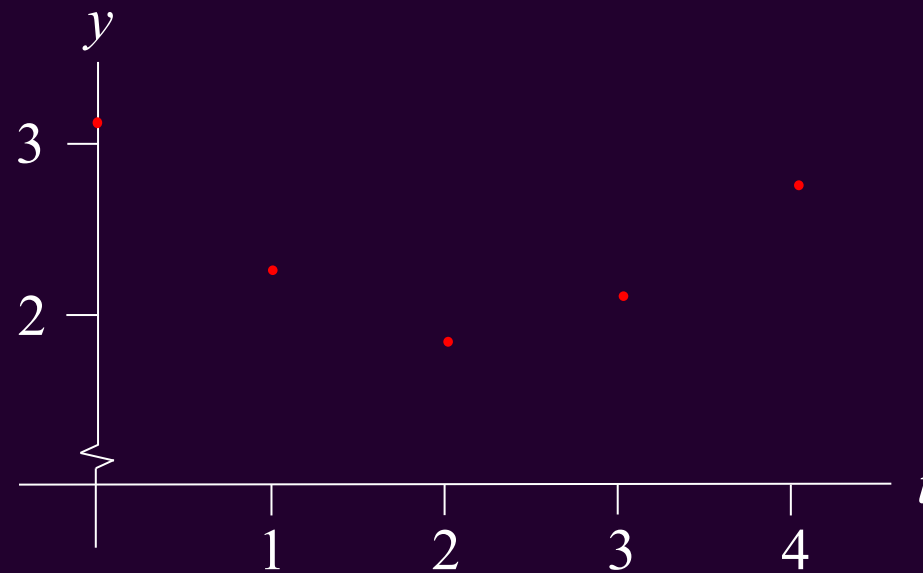




Applications

- Consider this data:

$(0, 3.1), (1, 2.2), (2, 1.9), (3, 2.1), (4, 2.8)$





Example

- Thus, we have these pairs of t - y points:

$$(0, 3.1), (1, 2.2), (2, 1.9), (3, 2.1), (4, 2.8)$$

- Suppose we want to find the best fitting quadratic $y = at^2 + bt + c$
 - Thus, we have five equations in three unknowns:

$$a \begin{pmatrix} 0 \\ 1 \\ 4 \\ 9 \\ 16 \end{pmatrix} + b \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} + c \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \approx \begin{pmatrix} 3.1 \\ 2.2 \\ 1.9 \\ 2.1 \\ 2.8 \end{pmatrix}$$

$$0a + 0b + c = 3.1$$

$$1a + 1b + c = 2.2$$

$$4a + 2b + c = 1.9$$

$$9a + 3b + c = 2.1$$

$$16a + 4b + c = 2.8$$

$$\mathbf{y} = \begin{pmatrix} 3.1 \\ 2.2 \\ 1.9 \\ 2.1 \\ 2.8 \end{pmatrix}$$





Example

- Thus, we have these pairs of t - y points:

$(0, 3.1), (1, 2.1), (2, 1.9), (3, 2.1), (4, 2.9)$

- Therefore, we must define

$$V = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \\ 16 & 4 & 1 \end{pmatrix} \quad \begin{array}{l} 0a + 0b + c = 3.1 \\ 1a + 1b + c = 2.2 \\ 4a + 2b + c = 1.9 \\ 9a + 3b + c = 2.1 \\ 16a + 4b + c = 2.9 \end{array} \quad \mathbf{y} = \begin{pmatrix} 3.1 \\ 2.2 \\ 1.9 \\ 2.1 \\ 2.8 \end{pmatrix}$$





Example

- Next we calculate:

$$V^T V = \begin{pmatrix} 0 & 1 & 4 & 9 & 16 \\ 0 & 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \\ 16 & 4 & 1 \end{pmatrix} = \begin{pmatrix} 354 & 100 & 30 \\ 100 & 30 & 10 \\ 30 & 10 & 5 \end{pmatrix}$$

$$\begin{pmatrix} 354 & 100 & 30 \\ 100 & 30 & 10 \\ 30 & 10 & 5 \end{pmatrix} \boldsymbol{\alpha} = \begin{pmatrix} 73.5 \\ 23.5 \\ 12.1 \end{pmatrix} \quad V^T \mathbf{y} = \begin{pmatrix} 0 & 1 & 4 & 9 & 16 \\ 0 & 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 3.1 \\ 2.2 \\ 1.9 \\ 2.1 \\ 2.8 \end{pmatrix} = \begin{pmatrix} 73.5 \\ 23.5 \\ 12.1 \end{pmatrix}$$
$$\boldsymbol{\alpha} = \begin{pmatrix} 0.26429 \\ -1.12714 \\ 3.08857 \end{pmatrix}$$

- Thus, the best-fitting line is $y = 0.26429t^2 - 1.12714t + 3.08857$

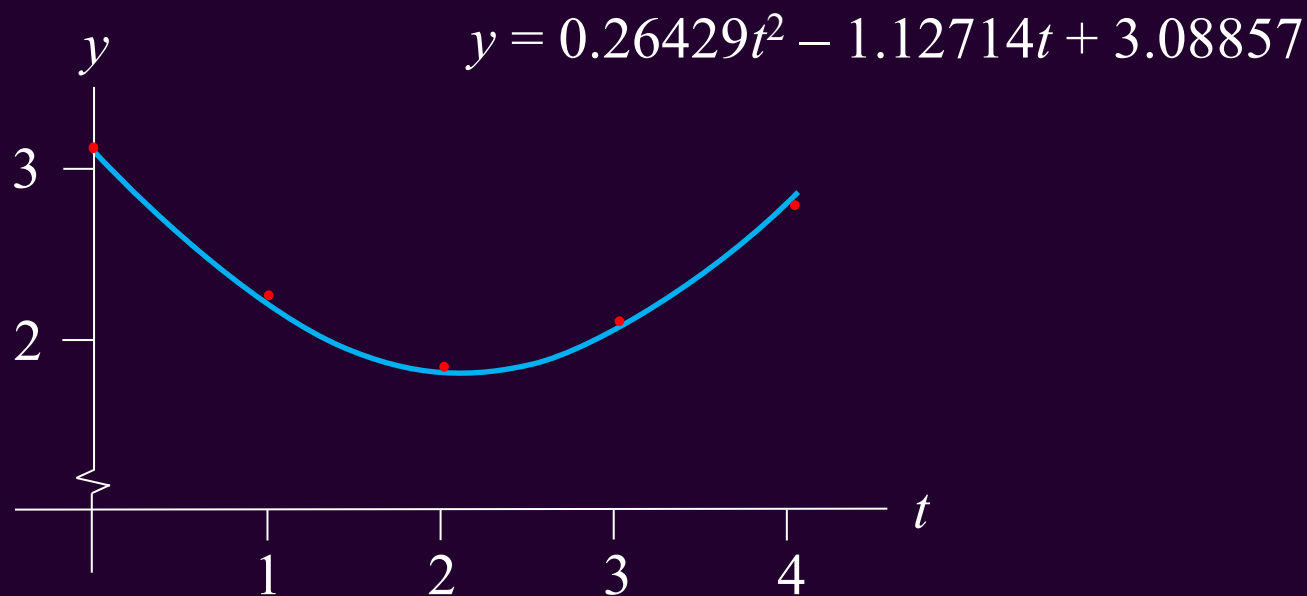




Applications

- Thus, looking at this result:

$(0, 3.1), (1, 2.2), (2, 1.9), (3, 2.1), (4, 2.8)$





Summary

- Following this topic, you now
 - Understand the idea of finding a best-fitting linear combination of vectors
 - Know that this results in a set of equations called the *normal equations*
 - Understand that if the column vectors are linearly independent, then V^*V has a solution
 - Looked at two applications of this technique to find a best-fitting line and best-fitting quadratic through data





References

- [1] https://en.wikipedia.org/wiki/Least_squares





Acknowledgments

None so far.





Colophon

These slides were prepared using the Cambria typeface. Mathematical equations use Times New Roman, and source code is presented using Consolas. Mathematical equations are prepared in MathType by Design Science, Inc. Examples may be formulated and checked using Maple by Maplesoft, Inc.

The photographs of flowers and a monarch butter appearing on the title slide and accenting the top of each other slide were taken at the Royal Botanical Gardens in October of 2017 by Douglas Wilhelm Harder. Please see

<https://www.rbg.ca/>

for more information.





Disclaimer

These slides are provided for the NE 112 *Linear algebra for nanotechnology engineering* course taught at the University of Waterloo. The material in it reflects the authors' best judgment in light of the information available to them at the time of preparation. Any reliance on these course slides by any party for any other purpose are the responsibility of such parties. The authors accept no responsibility for damages, if any, suffered by any party as a result of decisions made or actions based on these course slides for any other purpose than that for which it was intended.

