Introduction to Probabilistic Machine Learning

Ralf Herbrich, Rainer Schlosser

Hasso Plattner Institut

Tutorial 7 – Recap Theory Unit 7

Overview

1. **Questions and Updates**
2. Recap: Unit 7, Linear Algebra, and the Inverse of a Matrix
3. Recap: Cholesky Decomposition
4. Recap: Linear Regression in Matrix Notation

Course Overview

Week	Topic Lecture	Tutorial	Exercises
07.04. & 08.04.	1 Probability Theory	Intro Julia	
14.04. & 15.04.	2 Inference Methods and Decision-Making	no tutorial	Exercise 1
21.04. & 22.04.	**no lecture**	Theory Unit 1 & 2	(14.04. – 08.05.)
28.04. & 29.04.	3 Graphical Models: Independence	Theory Unit 3	
05.05. & 06.05.	4 Graphical Models: Exact Inference	Theory Unit 4	Exercise 2
12.05. & 13.05.	5 Graphical Models: Approximate Inference	Theory Unit 5	(05.05. – 19.05.)
19.05. & 20.05.	6 Bayesian Ranking	Theory Unit 6	**Exercise 3**
26.05. & 27.05.	**7 Linear Basis Function Models**	**Theory Unit 7**	**(19.05. – 05.06.)**
02.06. & 03.06.	8 Bayesian Regression	Theory Unit 8	Exercise 4
09.06. & 10.06.	**no lecture**	9 Bayesian Classification	(02.06. – 23.06.)
16.06. & 17.06.	10 Non-Bayesian Classification Learning	Theory Unit 9 & 10	
23.06. & 24.06.	11 Gaussian Processes	Theory Unit 11	Exercise 5
30.06. & 01.07.	12 Information Theory	Theory Unit 12	(23.06. – 07.07.)
07.07. & 08.07.	13 Real-World Applications		

Overview

1. Questions and Updates
2. **Recap: Unit 7, Linear Algebra, and the Inverse of a Matrix**
3. Recap: Cholesky Decomposition
4. Recap: Linear Regression in Matrix Notation

Recap Unit 7: Overview of Concepts and Focus

a) Linear Basis Function Models

b) Modelling Data (Text and Images)

c) Linear Mappings and Matrices

d) Singular Value Decomposition

Matrix-Multiplication

Vector: An n-dimensional vector ***x*** is a column $\vec{x} = \begin{pmatrix} x_1 \\ \dots \\ x_n \end{pmatrix}$, $\vec{x}^T = \begin{pmatrix} x_1 & \dots & x_n \end{pmatrix}$ is a row

Matrix: An *n*x*m* matrix has *n* rows and *m* columns $\begin{pmatrix} m_{1,1} & & m_{1,m} \\ & & \\ m_{n,1} & & m_{n,m} \end{pmatrix}$

Matrix-Multiplication

Vector: An n-dimensional vector ***x*** is a column $\vec{x} = \begin{pmatrix} x_1 \\ \dots \\ x_n \end{pmatrix}$, $\vec{x}^T = \begin{pmatrix} x_1 & \dots & x_n \end{pmatrix}$ is a row

Matrix: An *n*x*m* matrix has *n* rows and *m* columns $\begin{pmatrix} m_{1,1} & m_{1,m} \\ & \\ m_{n,1} & m_{n,m} \end{pmatrix}$

Product: An *n*x*m* matrix times an *m*x*k* matrix is of *???* Format

Does the order matter? Is AxB = BxA ?

Matrix-Multiplication

Vector: An n-dimensional vector ***x*** is a column $\vec{x} = \begin{pmatrix} x_1 \\ \dots \\ x_n \end{pmatrix}$, $\vec{x}^T = \begin{pmatrix} x_1 & \dots & x_n \end{pmatrix}$ is a row

Matrix: An *n*x*m* matrix has *n* rows and *m* columns $\begin{pmatrix} m_{1,1} & & m_{1,m} \\ & & \\ m_{n,1} & & m_{n,m} \end{pmatrix}$

Product: An ***n***x*m* matrix times an *m*x***k*** matrix is of ***n***x***k*** format

Examples: $\underbrace{\begin{pmatrix} x_1 & \dots & x_n \end{pmatrix}}_{1xn} \cdot \underbrace{\begin{pmatrix} y_1 \\ \dots \\ y_n \end{pmatrix}}_{nx1} = \underbrace{(\dots)}_{1x1} \in \mathbb{R}$ $\quad$ $\underbrace{\begin{pmatrix} x_1 \\ \dots \\ x_n \end{pmatrix}}_{n\times 1} \cdot \underbrace{\begin{pmatrix} y_1 & \dots & y_n \end{pmatrix}}_{1\times n} = \underbrace{\begin{pmatrix} m_{1,1} & & m_{1,n} \\ & & \\ m_{n,1} & & m_{n,n} \end{pmatrix}}_{n\times n} \in \mathbb{R}^{n\times n}$

Recap: Inverse of a Matrix

Existence: An nxn matrix *A* is regular, i.e., has an inverse A^{-1} if, e.g.,

- *A* has full rank

- det(*A*)<>0

Property: $$A \cdot A^{-1} = A^{-1} \cdot A = I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & ... & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Solves: $$A \cdot \vec{x} = \vec{b} \quad \Leftrightarrow \quad \underbrace{A^{-1} \cdot A}_{I} \cdot \vec{x} = A^{-1} \cdot \vec{b}$$

To Solve Ax=b: Compute the Inverse of a Matrix

Computation 1: Given a regular nxn matrix *A*, the inverse A^{-1} can be computed as follows ??

To Solve Ax=b: Compute the Inverse of a Matrix

Computation 1: Given a regular nxn matrix *A*, the inverse A^{-1} can be computed as follows:

$$
\left(\begin{array}{c|ccc} & 1 & 0 & 0 \\ A & 0 & \dots & 0 \\ & 0 & 0 & 1 \end{array}\right)
$$

$$
\Leftrightarrow \qquad \dots
$$

$$
\Leftrightarrow \left(\begin{array}{ccc|c} 1 & 0 & 0 & \\ 0 & \dots & 0 & A^{-1} \\ 0 & 0 & 1 & \end{array}\right)
$$

Allowed: (i) switch rows, (ii) multiply row with nonzero scalars, (iii) add rows

To Solve Ax=b: Compute the Inverse of a Matrix

Computation 2: Given a regular nxn matrix *A*, the inverse A^{-1} can be computed as follows:

$$\begin{pmatrix} & & \\ & A & \\ & & \end{pmatrix} \cdot \begin{pmatrix} & & \\ & X & \\ & & \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \dots & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Solve a **Linear Program** (LP) with ??

To Solve Ax=b: Compute the Inverse of a Matrix

Computation 2: Given a regular nxn matrix *A*, the inverse A^{-1} can be computed as follows:

$$\begin{pmatrix} & & \\ & A & \\ & & \end{pmatrix} \cdot \begin{pmatrix} & & \\ & X & \\ & & \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \ldots & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Solve a **Linear Program** (LP) with nxn variables (cf. *X*) and nxn constraints:

$$\sum_{k=1,\ldots,n} A_{i,k} \cdot X_{k,j} = 1_{\{i=j\}} \qquad \forall i, j = 1,\ldots,n$$

To Solve Ax=b: Compute the Inverse of a Matrix

Computation 3: Given a regular nxn matrix *A*, the inverse A^{-1} can be computed as follows:

Use **Singular Value Decomposition** (U, V orthogonal, Σ diagonal):

$$A = U \cdot \Sigma \cdot V^T \quad \Leftrightarrow \quad A^{-1} = ??$$

To Solve Ax=b: Compute the Inverse of a Matrix

Computation 3: Given a regular nxn matrix *A*, the inverse A^{-1} can be computed as follows:

Use **Singular Value Decomposition** (U, V orthogonal, Σ diagonal):

$$A = U \cdot \Sigma \cdot V^T \quad \Leftrightarrow \quad A^{-1} = V \cdot \Sigma^{-1} \cdot U^T$$

Note:

$$V^T \cdot V = U^T \cdot U = I$$

$$A \cdot A^{-1} = U \cdot \underbrace{\Sigma \cdot \underbrace{V^T \cdot V}_{I} \cdot \Sigma^{-1}}_{I} \cdot U^T = I$$

Overview

1. Questions and Updates
2. Recap: Unit 7, Linear Algebra, and the Inverse of a Matrix
3. **Recap: Cholesky Decomposition**
4. Recap: Linear Regression in Matrix Notation

To Solve Ax=b: Cholesky Decomposition

Computation 4: Given a nxn matrix *A* (**symmetric**, positive-semidefinite)

we can compute a lower triangular matrix *L* such that:

$$L \cdot L^T = \begin{pmatrix} . & 0 & 0 \\ \bullet & . & 0 \\ \bullet & \bullet & . \end{pmatrix} \cdot \begin{pmatrix} . & \bullet & \bullet \\ 0 & . & \bullet \\ 0 & 0 & . \end{pmatrix} = \begin{pmatrix} & & \\ & A & \\ & & \end{pmatrix}$$

Then (Outlook): $Ansatz: \; L \cdot \underbrace{L^T \vec{x}}_{\vec{y}} = \vec{b} \qquad Then: \; L, \vec{b} \Rightarrow \vec{y} \qquad L, \vec{y} \Rightarrow \vec{x}$ (fast & stable!)

How to get *L*? Note, again the ~nxn/2 constraints determine *L*:

$$\sum_{k=1,\ldots,n} L_{i,k} \cdot L^T_{k,j} = \sum_{k=1,\ldots,n} L_{i,k} \cdot L_{j,k} = A_{i,j} \qquad \forall i, j = 1, \ldots, n$$

To Solve Ax=b: Cholesky Decomposition

$$
\begin{aligned}
\mathbf{A} = \mathbf{L}\mathbf{L}^T &= \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} L_{11} & L_{21} & L_{31} \\ 0 & L_{22} & L_{32} \\ 0 & 0 & L_{33} \end{pmatrix} \\
&= \begin{pmatrix} L_{11}^2 & & \text{(symmetric)} \\ L_{21}L_{11} & L_{21}^2 + L_{22}^2 & \\ L_{31}L_{11} & L_{31}L_{21} + L_{32}L_{22} & L_{31}^2 + L_{32}^2 + L_{33}^2 \end{pmatrix}
\end{aligned}
$$

Tutorial 7

PML SS 2025

To Solve Ax=b: Cholesky Decomposition

$$\mathbf{A} = \mathbf{L}\mathbf{L}^T = \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} L_{11} & L_{21} & L_{31} \\ 0 & L_{22} & L_{32} \\ 0 & 0 & L_{33} \end{pmatrix}$$

$$= \begin{pmatrix} \boxed{L_{11}^2} & & \text{(symmetric)} \\ L_{21}L_{11} & L_{21}^2 + L_{22}^2 & \\ L_{31}L_{11} & L_{31}L_{21} + L_{32}L_{22} & L_{31}^2 + L_{32}^2 + L_{33}^2 \end{pmatrix}$$

$$\sum_{k=1,\dots,n} L_{1,k} \cdot L_{1,k} = L_{1,1} \cdot L_{1,1} + 0 \cdot 0 + 0 \cdot 0 = A_{1,1}$$

To Solve Ax=b: Cholesky Decomposition

$$\mathbf{A} = \mathbf{L}\mathbf{L}^T = \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} L_{11} & L_{21} & L_{31} \\ 0 & L_{22} & L_{32} \\ 0 & 0 & L_{33} \end{pmatrix}$$

$$= \begin{pmatrix} \boxed{L_{11}^2} & & \text{(symmetric)} \\ \boxed{L_{21}L_{11}} & L_{21}^2 + L_{22}^2 & \\ L_{31}L_{11} & L_{31}L_{21} + L_{32}L_{22} & L_{31}^2 + L_{32}^2 + L_{33}^2 \end{pmatrix}$$

$$\sum_{k=1,\ldots,n} L_{1,k} \cdot L_{1,k} = L_{1,1} \cdot L_{1,1} + 0 \cdot 0 + 0 \cdot 0 = A_{1,1}$$

$$\sum_{k=1,\ldots,n} L_{2,k} \cdot L_{1,k} = L_{2,1} \cdot L_{1,1} + L_{2,2} \cdot 0 + 0 \cdot 0 = A_{2,1}$$

To Solve Ax=b: Cholesky Decomposition

$$\mathbf{A} = \mathbf{L}\mathbf{L}^T = \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} L_{11} & L_{21} & L_{31} \\ 0 & L_{22} & L_{32} \\ 0 & 0 & L_{33} \end{pmatrix}$$

$$= \begin{pmatrix} \boxed{L_{11}^2} & & \text{(symmetric)} \\ \boxed{L_{21}L_{11}} & \boxed{L_{21}^2 + L_{22}^2} & \\ L_{31}L_{11} & L_{31}L_{21} + L_{32}L_{22} & L_{31}^2 + L_{32}^2 + L_{33}^2 \end{pmatrix}$$

$$\sum_{k=1,\ldots,n} L_{1,k} \cdot L_{1,k} = L_{1,1} \cdot L_{1,1} + 0 \cdot 0 + 0 \cdot 0 = A_{1,1}$$

$$\sum_{k=1,\ldots,n} L_{2,k} \cdot L_{1,k} = L_{2,1} \cdot L_{1,1} + L_{2,2} \cdot 0 + 0 \cdot 0 = A_{2,1}$$

$$\sum_{k=1,\ldots,n} L_{2,k} \cdot L_{2,k} = L_{2,1} \cdot L_{2,1} + L_{2,2} \cdot L_{2,2} + 0 \cdot 0 = A_{2,2}$$

$$\ldots$$

To Solve Ax=b: Cholesky Decomposition

$$\mathbf{A} = \mathbf{L}\mathbf{L}^T = \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} L_{11} & L_{21} & L_{31} \\ 0 & L_{22} & L_{32} \\ 0 & 0 & L_{33} \end{pmatrix}$$

$$= \begin{pmatrix} \boxed{L_{11}^2} & & \text{(symmetric)} \\ \boxed{L_{21}L_{11}} & \boxed{L_{21}^2 + L_{22}^2} & \\ L_{31}L_{11} & L_{31}L_{21} + L_{32}L_{22} & L_{31}^2 + L_{32}^2 + L_{33}^2 \end{pmatrix}$$

$$\sum_{k=1,\dots,n} L_{1,k} \cdot L_{1,k} = L_{1,1} \cdot L_{1,1} + 0 \cdot 0 + 0 \cdot 0 = A_{1,1} \longrightarrow L_{1,1} \cdot L_{1,1} = A_{1,1} \Rightarrow L_{1,1} = \sqrt{A_{1,1}}$$

$$\sum_{k=1,\dots,n} L_{2,k} \cdot L_{1,k} = L_{2,1} \cdot L_{1,1} + L_{2,2} \cdot 0 + 0 \cdot 0 = A_{2,1}$$

$$\sum_{k=1,\dots,n} L_{2,k} \cdot L_{2,k} = L_{2,1} \cdot L_{2,1} + L_{2,2} \cdot L_{2,2} + 0 \cdot 0 = A_{2,2}$$

$$\dots$$

To Solve Ax=b: Cholesky Decomposition

$$\mathbf{A} = \mathbf{L}\mathbf{L}^T = \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} L_{11} & L_{21} & L_{31} \\ 0 & L_{22} & L_{32} \\ 0 & 0 & L_{33} \end{pmatrix}$$

$$= \begin{pmatrix} \boxed{L_{11}^2} & & \text{(symmetric)} \\ \boxed{L_{21}L_{11}} & \boxed{L_{21}^2 + L_{22}^2} & \\ L_{31}L_{11} & L_{31}L_{21} + L_{32}L_{22} & L_{31}^2 + L_{32}^2 + L_{33}^2 \end{pmatrix}$$

$$\sum_{k=1,\dots,n} L_{1,k} \cdot L_{1,k} = L_{1,1} \cdot L_{1,1} + 0 \cdot 0 + 0 \cdot 0 = A_{1,1} \longrightarrow L_{1,1} \cdot L_{1,1} = A_{1,1} \Rightarrow L_{1,1} = \sqrt{A_{1,1}}$$

$$\sum_{k=1,\dots,n} L_{2,k} \cdot L_{1,k} = L_{2,1} \cdot L_{1,1} + L_{2,2} \cdot 0 + 0 \cdot 0 = A_{2,1} \longrightarrow L_{2,1} \cdot L_{1,1} = A_{2,1} \Rightarrow L_{2,1} = A_{2,1} / L_{1,1} = A_{2,1} / \sqrt{A_{1,1}}$$

$$\sum_{k=1,\dots,n} L_{2,k} \cdot L_{2,k} = L_{2,1} \cdot L_{2,1} + L_{2,2} \cdot L_{2,2} + 0 \cdot 0 = A_{2,2}$$

...

To Solve Ax=b: Cholesky Decomposition

$$\mathbf{A} = \mathbf{L}\mathbf{L}^T = \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} L_{11} & L_{21} & L_{31} \\ 0 & L_{22} & L_{32} \\ 0 & 0 & L_{33} \end{pmatrix}$$

$$= \begin{pmatrix} \boxed{L_{11}^2} & & \text{(symmetric)} \\ \boxed{L_{21}L_{11}} & \boxed{L_{21}^2 + L_{22}^2} & \\ L_{31}L_{11} & L_{31}L_{21} + L_{32}L_{22} & L_{31}^2 + L_{32}^2 + L_{33}^2 \end{pmatrix}$$

$$\sum_{k=1,\dots,n} L_{1,k} \cdot L_{1,k} = L_{1,1} \cdot L_{1,1} + 0 \cdot 0 + 0 \cdot 0 = A_{1,1} \longrightarrow L_{1,1} \cdot L_{1,1} = A_{1,1} \Rightarrow L_{1,1} = \sqrt{A_{1,1}}$$

$$\sum_{k=1,\dots,n} L_{2,k} \cdot L_{1,k} = L_{2,1} \cdot L_{1,1} + L_{2,2} \cdot 0 + 0 \cdot 0 = A_{2,1} \longrightarrow L_{2,1} \cdot L_{1,1} = A_{2,1} \Rightarrow L_{2,1} = A_{2,1} / L_{1,1} = A_{2,1} / \sqrt{A_{1,1}}$$

$$\sum_{k=1,\dots,n} L_{2,k} \cdot L_{2,k} = L_{2,1} \cdot L_{2,1} + L_{2,2} \cdot L_{2,2} + 0 \cdot 0 = A_{2,2} \longrightarrow L_{2,1}^2 \cdot L_{2,2}^2 = A_{2,2} \Rightarrow L_{2,2} = \sqrt{A_{2,2} / L_{2,1}^2}$$

...

To Solve Ax=b: Cholesky Decomposition

$$\mathbf{A} = \mathbf{L}\mathbf{L}^T = \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} L_{11} & L_{21} & L_{31} \\ 0 & L_{22} & L_{32} \\ 0 & 0 & L_{33} \end{pmatrix}$$

$$= \begin{pmatrix} \boxed{L_{11}^2} & & \text{(symmetric)} \\ \boxed{L_{21}} L_{11} & L_{21}^2 + \boxed{L_{22}^2} & \\ L_{31} L_{11} & L_{31} L_{21} + L_{32} L_{22} & L_{31}^2 + L_{32}^2 + L_{33}^2 \end{pmatrix}$$

$$\mathbf{L} = \begin{pmatrix} \sqrt{A_{11}} & 0 & 0 \\ A_{21}/L_{11} & \sqrt{A_{22} - L_{21}^2} & 0 \\ A_{31}/L_{11} & \left(A_{32} - L_{31} L_{21}\right)/L_{22} & \sqrt{A_{33} - L_{31}^2 - L_{32}^2} \end{pmatrix}$$

Tutorial 7

PML SS 2025

To Solve Ax=b: Cholesky Decomposition

$$\mathbf{A} = \mathbf{L}\mathbf{L}^T = \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} L_{11} & L_{21} & L_{31} \\ 0 & L_{22} & L_{32} \\ 0 & 0 & L_{33} \end{pmatrix}$$

$$= \begin{pmatrix} \boxed{L_{11}^2} & & \text{(symmetric)} \\ \boxed{L_{21}} L_{11} & L_{21}^2 + \boxed{L_{22}^2} & \\ \boxed{L_{31}} L_{11} & L_{31}L_{21} + L_{32}L_{22} & L_{31}^2 + L_{32}^2 + L_{33}^2 \end{pmatrix}$$

$$\mathbf{L} = \begin{pmatrix} \sqrt{A_{11}} & 0 & 0 \\ A_{21}/L_{11} & \sqrt{A_{22} - L_{21}^2} & 0 \\ A_{31}/L_{11} & \left(A_{32} - L_{31}L_{21}\right)/L_{22} & \sqrt{A_{33} - L_{31}^2 - L_{32}^2} \end{pmatrix}$$

To Solve Ax=b: Cholesky Decomposition

$$\mathbf{A} = \mathbf{L}\mathbf{L}^T = \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} L_{11} & L_{21} & L_{31} \\ 0 & L_{22} & L_{32} \\ 0 & 0 & L_{33} \end{pmatrix}$$

$$= \begin{pmatrix} \boxed{L_{11}^2} & & \text{(symmetric)} \\ \boxed{L_{21}} L_{11} & L_{21}^2 + \boxed{L_{22}^2} & \\ \boxed{L_{31}} L_{11} & L_{31}L_{21} + \boxed{L_{32}} L_{22} & L_{31}^2 + L_{32}^2 + L_{33}^2 \end{pmatrix}$$

$$\mathbf{L} = \begin{pmatrix} \sqrt{A_{11}} & 0 & 0 \\ A_{21}/L_{11} & \sqrt{A_{22} - L_{21}^2} & 0 \\ A_{31}/L_{11} & \left(A_{32} - L_{31}L_{21}\right)/L_{22} & \sqrt{A_{33} - L_{31}^2 - L_{32}^2} \end{pmatrix}$$

To Solve Ax=b: Cholesky Decomposition

$$\mathbf{A} = \mathbf{L}\mathbf{L}^T = \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} L_{11} & L_{21} & L_{31} \\ 0 & L_{22} & L_{32} \\ 0 & 0 & L_{33} \end{pmatrix}$$

$$= \begin{pmatrix} \boxed{L_{11}^2} & & \text{(symmetric)} \\ \boxed{L_{21}} L_{11} & L_{21}^2 + \boxed{L_{22}^2} & \\ \boxed{L_{31}} L_{11} & L_{31} L_{21} + \boxed{L_{32}} L_{22} & L_{31}^2 + L_{32}^2 + \boxed{L_{33}^2} \end{pmatrix}$$

$$\mathbf{L} = \begin{pmatrix} \sqrt{A_{11}} & 0 & 0 \\ A_{21}/L_{11} & \sqrt{A_{22} - L_{21}^2} & 0 \\ A_{31}/L_{11} & \left(A_{32} - L_{31} L_{21}\right)/L_{22} & \sqrt{A_{33} - L_{31}^2 - L_{32}^2} \end{pmatrix}$$

To Solve Ax=b: Cholesky Decomposition

$$\mathbf{A} = \mathbf{L}\mathbf{L}^T = \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} L_{11} & L_{21} & L_{31} \\ 0 & L_{22} & L_{32} \\ 0 & 0 & L_{33} \end{pmatrix}$$

$$= \begin{pmatrix} \boxed{L_{11}^2} & & \text{(symmetric)} \\ \boxed{L_{21}} L_{11} & L_{21}^2 + \boxed{L_{22}^2} & \\ \boxed{L_{31}} L_{11} & L_{31} L_{21} + \boxed{L_{32}} L_{22} & L_{31}^2 + L_{32}^2 + \boxed{L_{33}^2} \end{pmatrix}$$

$$\mathbf{L} = \begin{pmatrix} \sqrt{A_{11}} & 0 & 0 \\ A_{21}/L_{11} & \sqrt{A_{22} - L_{21}^2} & 0 \\ A_{31}/L_{11} & \left(A_{32} - L_{31} L_{21}\right)/L_{22} & \sqrt{A_{33} - L_{31}^2 - L_{32}^2} \end{pmatrix}$$

$$\textit{Order of computation} = \begin{pmatrix} 1 & & & \\ 2 & 3 & & \\ 2 & 4 & 5 & \\ 2 & 6 & 7 & 8 \end{pmatrix} \; \textit{or} \; \begin{pmatrix} 1 & & & \\ 2 & 3 & & \\ 4 & 5 & 6 & \\ 7 & 8 & 9 & 10 \end{pmatrix}$$

General formula to compute *L* ??

To Solve Ax=b: Cholesky Decomposition

$$\mathbf{A} = \mathbf{L}\mathbf{L}^T = \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} L_{11} & L_{21} & L_{31} \\ 0 & L_{22} & L_{32} \\ 0 & 0 & L_{33} \end{pmatrix}$$

$$= \begin{pmatrix} \boxed{L_{11}^2} & & (\text{symmetric}) \\ \boxed{L_{21}} L_{11} & L_{21}^2 + \boxed{L_{22}^2} & \\ \boxed{L_{31}} L_{11} & L_{31} L_{21} + \boxed{L_{32}} L_{22} & L_{31}^2 + L_{32}^2 + \boxed{L_{33}^2} \end{pmatrix}$$

$$\mathbf{L} = \begin{pmatrix} \sqrt{A_{11}} & 0 & 0 \\ A_{21}/L_{11} & \sqrt{A_{22} - L_{21}^2} & 0 \\ A_{31}/L_{11} & (A_{32} - L_{31} L_{21})/L_{22} & \sqrt{A_{33} - L_{31}^2 - L_{32}^2} \end{pmatrix}$$

$$\textit{Order of computation} = \begin{pmatrix} 1 & & & \\ 2 & 3 & & \\ 2 & 4 & 5 & \\ 2 & 6 & 7 & 8 \end{pmatrix} \; or \; \begin{pmatrix} 1 & & & \\ 2 & 3 & & \\ 4 & 5 & 6 & \\ 7 & 8 & 9 & 10 \end{pmatrix}$$

General formula to compute L:

$$L_{i,j} = \begin{cases} \sqrt{A_{1,1}} & j = 1,\; i = 1 \\ A_{i,1} / L_{1,1} & j = 1,\; i > 1 \\ 0 & i < j \\ \sqrt{A_{i,j} - \sum\limits_{k=1,\dots,j-1} L_{j,k}^2} & i = j \\ \dfrac{1}{L_{j,j}} \cdot \left(A_{i,j} - \sum\limits_{k=1,\dots,j-1} L_{i,k} \cdot L_{j,k} \right) & i > j \end{cases}$$

Overview

1. Questions and Updates
2. Recap: Unit 7, Linear Algebra, and the Inverse of a Matrix
3. Recap: Cholesky Decomposition
4. **Recap: Linear Regression in Matrix Notation**

Recap: Regression & Linear Basis Function Models

Dependent variable y

Explanatory variables $x_0, x_1, \ldots, x_K$

Model: Find weights such that $\beta_0 \cdot x_0 + \beta_1 \cdot x_1 + \ldots + \beta_K \cdot x_K = \sum_{k=0,1,\ldots,K} \beta_k \cdot x_k \approx y$

Recap: Regression & Linear Basis Function Models

Dependent variable y

Explanatory variables $x_0, x_1, \ldots, x_K$

Model: Find weights such that $\beta_0 \cdot x_0 + \beta_1 \cdot x_1 + \ldots + \beta_K \cdot x_K = \sum_{k=0,1,\ldots,K} \beta_k \cdot x_k \approx y$

Data (Example): $y_i,\ z_i = (a_i, b_i) \qquad i = 1, \ldots, n$

Basis Functions: (i) $y_i \approx \beta_0 + \beta_1 \cdot a_i + \beta_2 \cdot b_i + \beta_3 \cdot \sqrt{a_i \cdot b_i}$

(ii)

(iii)

Recap: Regression & Linear Basis Function Models

Dependent variable y

Explanatory variables $x_0, x_1, \ldots, x_K$

Model: Find weights such that $\beta_0 \cdot x_0 + \beta_1 \cdot x_1 + \ldots + \beta_K \cdot x_K = \sum_{k=0,1,\ldots,K} \beta_k \cdot x_k \approx y$

Data (Example): $y_i,\ z_i = (a_i, b_i) \quad i = 1, \ldots, n$

Basis Functions: (i) $y_i \approx \beta_0 + \beta_1 \cdot a_i + \beta_2 \cdot b_i + \beta_3 \cdot \sqrt{a_i \cdot b_i}$

(ii) $y_i \approx \beta_0 + \beta_1 \cdot a_i + \beta_2 \cdot a_i^2 + \ldots + \beta_K \cdot a_i^K$

(iii)

Recap: Regression & Linear Basis Function Models

Dependent variable y

Explanatory variables $x_0, x_1, \ldots, x_K$

Model: Find weights such that $\beta_0 \cdot x_0 + \beta_1 \cdot x_1 + \ldots + \beta_K \cdot x_K = \sum_{k=0,1,\ldots,K} \beta_k \cdot x_k \approx y$

Data (Example): $y_i,\ z_i = (a_i, b_i) \qquad i = 1, \ldots, n$

Basis Functions:

(i) $y_i \approx \beta_0 + \beta_1 \cdot a_i + \beta_2 \cdot b_i + \beta_3 \cdot \sqrt{a_i \cdot b_i}$

(ii) $y_i \approx \beta_0 + \beta_1 \cdot a_i + \beta_2 \cdot a_i^2 + \ldots + \beta_K \cdot a_i^K$

(iii)
$$y_i \approx \beta_0 + \beta_1 \cdot a_i + \beta_2 \cdot a_i^2 + \ldots + \beta_K \cdot a_i^K + \beta_{K+1} \cdot b_i + \beta_{K+2} \cdot b_i^2 + \ldots + \beta_{K+K} \cdot b_i^K$$

Recap: Regression & Linear Basis Function Models

Dependent variable y

Explanatory variables $x_0, x_1, \dots, x_K$

Model: Find weights such that $\beta_0 \cdot x_0 + \beta_1 \cdot x_1 + \dots + \beta_K \cdot x_K = \sum_{k=0,1,\dots,K} \beta_k \cdot x_k \approx y$

Data (Example): $y_i,\ z_i = (a_i, b_i) \qquad i = 1,\dots,n$

Basis Functions:

(i) $y_i \approx \beta_0 + \beta_1 \cdot a_i + \beta_2 \cdot b_i + \beta_3 \cdot \sqrt{a_i \cdot b_i} \longrightarrow \sum_{k=0,1,\dots,3} \beta_k \cdot x_k^{(i)} \approx y_i$

(ii) $y_i \approx \beta_0 + \beta_1 \cdot a_i + \beta_2 \cdot a_i^2 + \dots + \beta_K \cdot a_i^K \longrightarrow \sum_{k=0,1,\dots,K} \beta_k \cdot x_k^{(i)} \approx y_i$

(iii) $y_i \approx \beta_0 + \beta_1 \cdot a_i + \beta_2 \cdot a_i^2 + \dots + \beta_K \cdot a_i^K + \beta_{K+1} \cdot b_i + \beta_{K+2} \cdot b_i^2 + \dots + \beta_{K+K} \cdot b_i^K \longrightarrow \sum_{k=0,1,\dots,2K} \beta_k \cdot x_k^{(i)} \approx y_i$

Recap: Linear Regression

Dependent variable y

Explanatory variables $x_0, x_1, \ldots, x_K$ (cf. Basis function models)

Model: Find weights such that $\beta_0 \cdot x_0 + \beta_1 \cdot x_1 + \ldots + \beta_K \cdot x_K = \sum_{k=0,1,\ldots,K} \beta_k \cdot x_k \approx y$

Objective of the Regression? What are the variables?

Recap: Linear Regression

Dependent variable y

Explanatory variables $x_0, x_1, \ldots, x_K$ (cf. Basis function models)

Model: Find weights such that $\beta_0 \cdot x_0 + \beta_1 \cdot x_1 + \ldots + \beta_K \cdot x_K = \sum_{k=0,1,\ldots,K} \beta_k \cdot x_k \approx y$

Minimize the Sum of Squared Errors:

$$\min_{\beta_k \in \mathbb{R}, k=0,\ldots,K} \sum_{i=1,\ldots,N} \left(\beta_0 \cdot x_{0,i} + \beta_1 \cdot x_{1,i} + \ldots + \beta_K \cdot x_{K,i} - y_i\right)^2$$

$$\simeq \vec{\beta}^* := \underset{\vec{\beta} \in \mathbb{R}^{K+1}}{\arg\min} \sum_{i=1,\ldots,N} \underbrace{\left(y_i - \sum_{j=1,\ldots,k} \beta_j \cdot x_{j,i}\right)^2}_{errors\ \varepsilon_i}$$

Let`s take a Look: Residuals and Matrix Notation

Consider the errors: $\varepsilon_i := y_i - \sum_{k=0,1,\ldots,K} \beta_k \cdot x_{k,i} \qquad i = 1,\ldots,n$

What is happening here in Matrix Notation?

Let`s take a Look: Residuals and Matrix Notation

Consider the errors: $\varepsilon_i := y_i - \sum_{k=0,1,\ldots,K} \beta_k \cdot x_{k,i} \qquad i = 1,\ldots,n$

What is happening here in Matrix Notation?

We have:

$$\begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{pmatrix}_{n\times 1} = \begin{pmatrix} 1 & x_{1,1} & x_{2,1} & \ldots & x_{K,1} \\ 1 & x_{1,2} & x_{2,2} & & x_{K,2} \\ \ldots & & & \ldots & \\ 1 & x_{1,n} & x_{2,n} & & x_{K,n} \end{pmatrix}_{n\times(K+1)} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \ldots \\ \beta_K \end{pmatrix}_{(K+1)\times 1} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \ldots \\ \varepsilon_n \end{pmatrix}_{n\times 1}$$

Or just ??

Let‵s take a Look: Residuals and Matrix Notation

Consider the errors: $\varepsilon_i := y_i - \sum_{k=0,1,\ldots,K} \beta_k \cdot x_{k,i} \qquad i = 1,\ldots,n$

What is happening here in Matrix Notation?

We have:

$$\begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{pmatrix}_{n\times 1} = \begin{pmatrix} 1 & x_{1,1} & x_{2,1} & \ldots & x_{K,1} \\ 1 & x_{1,2} & x_{2,2} & & x_{K,2} \\ \ldots & & & \ldots & \\ 1 & x_{1,n} & x_{2,n} & & x_{K,n} \end{pmatrix}_{n\times(K+1)} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \ldots \\ \beta_K \end{pmatrix}_{(K+1)\times 1} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \ldots \\ \varepsilon_n \end{pmatrix}_{n\times 1}$$

Or just:

$$\vec{y} = X\vec{\beta} + \vec{\varepsilon} \quad \Leftrightarrow \quad \vec{\varepsilon} = \vec{y} - X\vec{\beta}$$

Sum of Squared Errors

Consider the errors: $\varepsilon_i := y_i - \sum_{k=0,1,\ldots,K} \beta_k \cdot x_{k,i} \qquad i = 1,\ldots,n \qquad or \qquad \vec{\varepsilon} = \vec{y} - X\vec{\beta}$

Sum of squared errors in Matrix notation?

Sum of Squared Errors

Consider the errors: $\varepsilon_i := y_i - \sum_{k=0,1,\ldots,K} \beta_k \cdot x_{k,i} \qquad i = 1,\ldots,n \qquad or \qquad \vec{\varepsilon} = \vec{y} - X\vec{\beta}$

Sum of squared errors:

$$\begin{pmatrix} \varepsilon_1 & \varepsilon_2 & \ldots & \varepsilon_n \end{pmatrix}_{1\times n} \cdot \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \ldots \\ \varepsilon_n \end{pmatrix}_{n\times 1} = \varepsilon_1^2 + \varepsilon_2^2 + \ldots + \varepsilon_n^2 = \vec{\varepsilon}^T \vec{\varepsilon}$$

Great! What now?

Sum of Squared Errors

Consider the errors: $\varepsilon_i := y_i - \sum_{k=0,1,\ldots,K} \beta_k \cdot x_{k,i} \qquad i = 1,\ldots,n \qquad or \qquad \vec{\varepsilon} = \vec{y} - X\vec{\beta}$

Sum of squared errors:

$$\begin{pmatrix} \varepsilon_1 & \varepsilon_2 & \ldots & \varepsilon_n \end{pmatrix}_{1\times n} \cdot \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \ldots \\ \varepsilon_n \end{pmatrix}_{n\times 1} = \varepsilon_1^2 + \varepsilon_2^2 + \ldots + \varepsilon_n^2 = \vec{\varepsilon}^T \vec{\varepsilon}$$

We have:

$$\vec{\varepsilon}^T \vec{\varepsilon} = \left(\vec{y} - X\vec{\beta}\right)^T \cdot \left(\vec{y} - X\vec{\beta}\right)$$
$$= ??$$

Sum of Squared Errors

Consider the errors: $\varepsilon_i := y_i - \sum_{k=0,1,\ldots,K} \beta_k \cdot x_{k,i} \qquad i = 1,\ldots,n \qquad or \qquad \vec{\varepsilon} = \vec{y} - X\vec{\beta}$

Sum of squared errors:

$$\begin{pmatrix} \varepsilon_1 & \varepsilon_2 & \ldots & \varepsilon_n \end{pmatrix}_{1\times n} \cdot \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \ldots \\ \varepsilon_n \end{pmatrix}_{n\times 1} = \varepsilon_1^2 + \varepsilon_2^2 + \ldots + \varepsilon_n^2 = \vec{\varepsilon}^T \vec{\varepsilon}$$

We have:

$$\begin{aligned} \vec{\varepsilon}^T \vec{\varepsilon} &= \left(\vec{y} - X\vec{\beta}\right)^T \cdot \left(\vec{y} - X\vec{\beta}\right) \\ &= \left(\vec{y}^T - (X\vec{\beta})^T\right) \cdot \left(\vec{y} - X\vec{\beta}\right) \\ &= ?? \end{aligned}$$

Sum of Squared Errors

Consider the errors: $\varepsilon_i := y_i - \sum_{k=0,1,\ldots,K} \beta_k \cdot x_{k,i} \qquad i = 1,\ldots,n \qquad or \qquad \vec{\varepsilon} = \vec{y} - X\vec{\beta}$

Sum of squared errors:

$$\begin{pmatrix} \varepsilon_1 & \varepsilon_2 & \ldots & \varepsilon_n \end{pmatrix}_{1\times n} \cdot \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \ldots \\ \varepsilon_n \end{pmatrix}_{n\times 1} = \varepsilon_1^2 + \varepsilon_2^2 + \ldots + \varepsilon_n^2 = \vec{\varepsilon}^T \vec{\varepsilon}$$

We have:

$$\begin{aligned} \vec{\varepsilon}^T \vec{\varepsilon} &= \left(\vec{y} - X\vec{\beta}\right)^T \cdot \left(\vec{y} - X\vec{\beta}\right) \\ &= \left(\vec{y}^T - (X\vec{\beta})^T\right) \cdot \left(\vec{y} - X\vec{\beta}\right) \\ &= \vec{y}^T \vec{y} - \vec{y}^T X\vec{\beta} - (X\vec{\beta})^T \vec{y} + (X\vec{\beta})^T X\vec{\beta} \end{aligned}$$

Tutorial 7

PML SS 2025

Transposed Terms

Consider:

$$X\vec{\beta} = \begin{pmatrix} 1 & x_{1,1} & x_{2,1} & \dots & x_{K,1} \\ 1 & x_{1,2} & x_{2,2} & & x_{K,2} \\ \dots & & & \dots & \\ 1 & x_{1,n} & x_{2,n} & & x_{K,n} \end{pmatrix}_{n\times(K+1)} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_K \end{pmatrix}_{(K+1)\times 1} = \begin{pmatrix} \vec{x}_1{}^T \vec{\beta} \\ \vec{x}_2{}^T \vec{\beta} \\ \dots \\ \vec{x}_n{}^T \vec{\beta} \end{pmatrix}_{n\times 1}$$

Then:

$$(X\vec{\beta})^T = ??$$

Tutorial 7

PML SS 2025

Transposed Terms

Consider:

$$X\vec{\beta} = \begin{pmatrix} 1 & x_{1,1} & x_{2,1} & \dots & x_{K,1} \\ 1 & x_{1,2} & x_{2,2} & & x_{K,2} \\ \dots & & & \dots & \\ 1 & x_{1,n} & x_{2,n} & & x_{K,n} \end{pmatrix}_{n\times(K+1)} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_K \end{pmatrix}_{(K+1)\times 1} = \begin{pmatrix} \vec{x}_1{}^T\vec{\beta} \\ \vec{x}_2{}^T\vec{\beta} \\ \dots \\ \vec{x}_n{}^T\vec{\beta} \end{pmatrix}_{n\times 1}$$

Then:

$$(X\vec{\beta})^T = \begin{pmatrix} \vec{x}_1{}^T\vec{\beta} & \vec{x}_2{}^T\vec{\beta} & \dots & \vec{x}_n{}^T\vec{\beta} \end{pmatrix}_{1\times n} = ??$$

Transposed Terms

Consider:
$$X\vec{\beta} = \begin{pmatrix} 1 & x_{1,1} & x_{2,1} & \dots & x_{K,1} \\ 1 & x_{1,2} & x_{2,2} & & x_{K,2} \\ \dots & & & \dots & \\ 1 & x_{1,n} & x_{2,n} & & x_{K,n} \end{pmatrix}_{n\times(K+1)} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_K \end{pmatrix}_{(K+1)\times 1} = \begin{pmatrix} \vec{x}_1^{\,T}\vec{\beta} \\ \vec{x}_2^{\,T}\vec{\beta} \\ \dots \\ \vec{x}_n^{\,T}\vec{\beta} \end{pmatrix}_{n\times 1}$$

We have:
$$(X\vec{\beta})^T = \begin{pmatrix} \vec{x}_1^{\,T}\vec{\beta} & \vec{x}_2^{\,T}\vec{\beta} & \dots & \vec{x}_n^{\,T}\vec{\beta} \end{pmatrix}_{1\times n} = \begin{pmatrix} \beta_0 & \beta_1 & \dots & \beta_K \end{pmatrix}_{1\times(K+1)} \cdot \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{1,1} & x_{1,2} & & x_{1,n} \\ x_{2,1} & x_{2,2} & & x_{2,n} \\ \dots & & \dots & \\ x_{K,1} & x_{K,2} & & x_{K,n} \end{pmatrix}_{(K+1)\times n}$$

Transposed Terms

Consider:

$$X\vec{\beta} = \begin{pmatrix} 1 & x_{1,1} & x_{2,1} & \dots & x_{K,1} \\ 1 & x_{1,2} & x_{2,2} & & x_{K,2} \\ \dots & & & \dots & \\ 1 & x_{1,n} & x_{2,n} & & x_{K,n} \end{pmatrix}_{n\times(K+1)} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_K \end{pmatrix}_{(K+1)\times 1} = \begin{pmatrix} \vec{x}_1{}^T\vec{\beta} \\ \vec{x}_2{}^T\vec{\beta} \\ \dots \\ \vec{x}_n{}^T\vec{\beta} \end{pmatrix}_{n\times 1}$$

We have:

$$(X\vec{\beta})^T = \left(\vec{x}_1{}^T\vec{\beta} \quad \vec{x}_2{}^T\vec{\beta} \quad \dots \quad \vec{x}_n{}^T\vec{\beta}\right)_{1\times n} = \left(\beta_0 \quad \beta_1 \quad \dots \quad \beta_K\right)_{1\times(K+1)} \cdot \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{1,1} & x_{1,2} & & x_{1,n} \\ x_{2,1} & x_{2,2} & & x_{2,n} \\ \dots & & \dots & \\ x_{K,1} & x_{K,2} & & x_{K,n} \end{pmatrix}_{(K+1)\times n}$$

$$= \vec{\beta}^T X^T$$

Note: Switch the order!

Sum of Squared Errors

Consider the errors: $\varepsilon_i := y_i - \sum_{k=0,1,\dots,K} \beta_k \cdot x_{k,i} \qquad i = 1,\dots,n \qquad or \qquad \vec{\varepsilon} = \vec{y} - X\vec{\beta}$

Sum of squared errors: $\varepsilon_1^2 + \varepsilon_2^2 + \dots + \varepsilon_n^2 = \vec{\varepsilon}^T \vec{\varepsilon}$

We have:

$$
\begin{aligned}
\vec{\varepsilon}^T \vec{\varepsilon} &= \left(\vec{y} - X\vec{\beta}\right)^T \cdot \left(\vec{y} - X\vec{\beta}\right) \\
&= \left(\vec{y}^T - (X\vec{\beta})^T\right) \cdot \left(\vec{y} - X\vec{\beta}\right) \\
&= \vec{y}^T \vec{y} - \vec{y}^T X\vec{\beta} - \boxed{(X\vec{\beta})^T \vec{y}} + (X\vec{\beta})^T X\vec{\beta} \\
&= \vec{y}^T \vec{y} - \vec{y}^T X\vec{\beta} - \boxed{\vec{\beta}^T X^T \vec{y}} + \vec{\beta}^T X^T X\vec{\beta} \\
&= ??
\end{aligned}
$$

Sum of Squared Errors

Consider the errors: $\varepsilon_i := y_i - \sum_{k=0,1,\dots,K} \beta_k \cdot x_{k,i} \qquad i = 1,\dots,n \qquad or \qquad \vec{\varepsilon} = \vec{y} - X\vec{\beta}$

Sum of squared errors: $\varepsilon_1^2 + \varepsilon_2^2 + \dots + \varepsilon_n^2 = \vec{\varepsilon}^T \vec{\varepsilon}$

We have:

$$\begin{aligned}
\vec{\varepsilon}^T \vec{\varepsilon} &= \left(\vec{y} - X\vec{\beta}\right)^T \cdot \left(\vec{y} - X\vec{\beta}\right) \\
&= \left(\vec{y}^T - (X\vec{\beta})^T\right) \cdot \left(\vec{y} - X\vec{\beta}\right) \\
&= \vec{y}^T\vec{y} - \vec{y}^T X\vec{\beta} - (X\vec{\beta})^T \vec{y} + (X\vec{\beta})^T X\vec{\beta} \\
&= \vec{y}^T\vec{y} - \vec{y}^T X\vec{\beta} - \vec{\beta}^T X^T \vec{y} + \vec{\beta}^T X^T X\vec{\beta} \\
&= \vec{y}^T\vec{y} - \underbrace{\left(\vec{y}^T X\vec{\beta}\right)^T}_{\in\mathbb{R}} - \vec{\beta}^T X^T \vec{y} + \vec{\beta}^T X^T X\vec{\beta} \\
&= ??
\end{aligned}$$

Sum of Squared Errors

Consider the errors: $\varepsilon_i := y_i - \sum_{k=0,1,\ldots,K} \beta_k \cdot x_{k,i} \qquad i = 1,\ldots,n \qquad or \qquad \vec{\varepsilon} = \vec{y} - X\vec{\beta}$

Sum of squared errors: $\varepsilon_1^2 + \varepsilon_2^2 + \ldots + \varepsilon_n^2 = \vec{\varepsilon}^T \vec{\varepsilon}$

We have:

$$
\begin{aligned}
\vec{\varepsilon}^T \vec{\varepsilon} &= \left(\vec{y} - X\vec{\beta}\right)^T \cdot \left(\vec{y} - X\vec{\beta}\right) \\
&= \left(\vec{y}^T - (X\vec{\beta})^T\right) \cdot \left(\vec{y} - X\vec{\beta}\right) \\
&= \vec{y}^T \vec{y} - \vec{y}^T X \vec{\beta} - (X\vec{\beta})^T \vec{y} + (X\vec{\beta})^T X \vec{\beta} \\
&= \vec{y}^T \vec{y} - \vec{y}^T X \vec{\beta} - \vec{\beta}^T X^T \vec{y} + \vec{\beta}^T X^T X \vec{\beta} \\
&= \vec{y}^T \vec{y} - \boxed{\left(\vec{y}^T X \vec{\beta}\right)^T} - \vec{\beta}^T X^T \vec{y} + \vec{\beta}^T X^T X \vec{\beta} \\
&= \vec{y}^T \vec{y} - \boxed{\vec{\beta}^T X^T \vec{y}} - \vec{\beta}^T X^T \vec{y} + \vec{\beta}^T X^T X \vec{\beta} \\
&= \vec{y}^T \vec{y} - 2\vec{\beta}^T X^T \vec{y} + \vec{\beta}^T X^T X \vec{\beta}
\end{aligned}
$$

Minimizing the Sum of Squared Errors

Sum of squared errors: $\vec{\varepsilon}^T\vec{\varepsilon} = \vec{y}^T\vec{y} - 2\vec{\beta}^T X^T\vec{y} + \vec{\beta}^T X^T X\vec{\beta} \in \mathbb{R}$

Derivatives/FOC??

Minimizing the Sum of Squared Errors

Sum of squared errors: $\vec{\varepsilon}^T \vec{\varepsilon} = \vec{y}^T \vec{y} - 2\vec{\beta}^T X^T \vec{y} + \vec{\beta}^T X^T X \vec{\beta} \in \mathbb{R}$

Derivatives/FOC: $\frac{\partial}{\partial \beta_k} \varepsilon_1^2 + \varepsilon_2^2 + \ldots + \varepsilon_n^2 \quad \forall k = 0,1,\ldots,K \quad \Rightarrow \begin{pmatrix} \frac{\partial}{\partial \beta_0} \vec{\varepsilon}^T \vec{\varepsilon} \\ \frac{\partial}{\partial \beta_1} \vec{\varepsilon}^T \vec{\varepsilon} \\ \ldots \\ \frac{\partial}{\partial \beta_K} \vec{\varepsilon}^T \vec{\varepsilon} \end{pmatrix}_{(K+1)\times 1} = \frac{\partial \vec{\varepsilon}^T \vec{\varepsilon}}{\partial \vec{\beta}} \overset{!}{=} \vec{0}$

Hence: $\frac{\partial \vec{\varepsilon}^T \vec{\varepsilon}}{\partial \vec{\beta}} = ??$

Minimizing the Sum of Squared Errors

Sum of squared errors: $\vec{\varepsilon}^T \vec{\varepsilon} = \vec{y}^T \vec{y} - 2\vec{\beta}^T X^T \vec{y} + \vec{\beta}^T X^T X \vec{\beta} \in \mathbb{R}$

Derivatives/FOC: $\frac{\partial}{\partial \beta_k} \varepsilon_1^2 + \varepsilon_2^2 + \ldots + \varepsilon_n^2 \quad \forall k = 0,1,\ldots,K \quad \Rightarrow \begin{pmatrix} \frac{\partial}{\partial \beta_0} \vec{\varepsilon}^T \vec{\varepsilon} \\ \frac{\partial}{\partial \beta_1} \vec{\varepsilon}^T \vec{\varepsilon} \\ \ldots \\ \frac{\partial}{\partial \beta_K} \vec{\varepsilon}^T \vec{\varepsilon} \end{pmatrix}_{(K+1)\times 1} = \frac{\partial \vec{\varepsilon}^T \vec{\varepsilon}}{\partial \vec{\beta}} \overset{!}{=} \vec{0}$

Hence:

$$\frac{\partial \vec{\varepsilon}^T \vec{\varepsilon}}{\partial \vec{\beta}} = \frac{\partial}{\partial \vec{\beta}} \left(\vec{y}^T \vec{y} - 2\vec{\beta}^T X^T \vec{y} + \vec{\beta}^T X^T X \vec{\beta} \right)$$

$$= \underbrace{\frac{\partial}{\partial \vec{\beta}} \vec{y}^T \vec{y}}_{??} - \frac{\partial}{\partial \vec{\beta}} 2\vec{\beta}^T X^T \vec{y} + \frac{\partial}{\partial \vec{\beta}} \vec{\beta}^T X^T X \vec{\beta}$$

Minimizing the Sum of Squared Errors

Sum of squared errors: $\vec{\varepsilon}^T\vec{\varepsilon} = \vec{y}^T\vec{y} - 2\vec{\beta}^T X^T \vec{y} + \vec{\beta}^T X^T X \vec{\beta} \in \mathbb{R}$

Derivatives/FOC: $\frac{\partial}{\partial \beta_k} \varepsilon_1^2 + \varepsilon_2^2 + \ldots + \varepsilon_n^2 \quad \forall k = 0,1,\ldots,K \Rightarrow \begin{pmatrix} \frac{\partial}{\partial \beta_0}\vec{\varepsilon}^T\vec{\varepsilon} \\ \frac{\partial}{\partial \beta_1}\vec{\varepsilon}^T\vec{\varepsilon} \\ \ldots \\ \frac{\partial}{\partial \beta_K}\vec{\varepsilon}^T\vec{\varepsilon} \end{pmatrix}_{(K+1)\times 1} = \frac{\partial \vec{\varepsilon}^T\vec{\varepsilon}}{\partial \vec{\beta}} \overset{!}{=} \vec{0}$

Hence:
$$\frac{\partial \vec{\varepsilon}^T\vec{\varepsilon}}{\partial \vec{\beta}} = \frac{\partial}{\partial \vec{\beta}}\left(\vec{y}^T\vec{y} - 2\vec{\beta}^T X^T \vec{y} + \vec{\beta}^T X^T X \vec{\beta}\right)$$
$$= \underbrace{\frac{\partial}{\partial \vec{\beta}}\vec{y}^T\vec{y}}_{\begin{pmatrix} 0 \\ 0 \\ \ldots \\ 0 \end{pmatrix}_{(K+1)\times 1}} - \frac{\partial}{\partial \vec{\beta}} 2\vec{\beta}^T X^T \vec{y} + \frac{\partial}{\partial \vec{\beta}} \vec{\beta}^T X^T X \vec{\beta}$$

$$\boxed{\frac{\partial}{\partial \beta_k}}\vec{y}^T\vec{y} = \frac{\partial}{\partial \beta_k}\left(y_1^2 + y_2^2 + \ldots + y_n^2\right) = 0 \quad \forall k = 0,1,\ldots,K$$

Minimizing the Sum of Squared Errors

Sum of squared errors: $\vec{\varepsilon}^T \vec{\varepsilon} = \vec{y}^T \vec{y} - 2\vec{\beta}^T X^T \vec{y} + \vec{\beta}^T X^T X \vec{\beta} \in \mathbb{R}$

Derivatives/FOC: $\frac{\partial}{\partial \beta_k} \varepsilon_1^2 + \varepsilon_2^2 + \ldots + \varepsilon_n^2 \quad \forall k = 0,1,\ldots,K \quad \Rightarrow \begin{pmatrix} \frac{\partial}{\partial \beta_0} \vec{\varepsilon}^T \vec{\varepsilon} \\ \frac{\partial}{\partial \beta_1} \vec{\varepsilon}^T \vec{\varepsilon} \\ \ldots \\ \frac{\partial}{\partial \beta_K} \vec{\varepsilon}^T \vec{\varepsilon} \end{pmatrix}_{(K+1)\times 1} = \frac{\partial \vec{\varepsilon}^T \vec{\varepsilon}}{\partial \vec{\beta}} \overset{!}{=} \vec{0}$

Hence:

$$\frac{\partial \vec{\varepsilon}^T \vec{\varepsilon}}{\partial \vec{\beta}} = \frac{\partial}{\partial \vec{\beta}} \left(\vec{y}^T \vec{y} - 2\vec{\beta}^T X^T \vec{y} + \vec{\beta}^T X^T X \vec{\beta} \right)$$

$$= \vec{0} - \underbrace{\frac{\partial}{\partial \vec{\beta}} 2\vec{\beta}^T X^T \vec{y}}_{??} + \frac{\partial}{\partial \vec{\beta}} \vec{\beta}^T X^T X \vec{\beta}$$

Tutorial 7

PML SS 2025

Minimizing the Sum of Squared Errors

Sum of squared errors: $\vec{\varepsilon}^T\vec{\varepsilon} = \vec{y}^T\vec{y} - 2\vec{\beta}^T X^T \vec{y} + \vec{\beta}^T X^T X \vec{\beta} \in \mathbb{R}$

Derivatives/FOC: $\frac{\partial}{\partial \beta_k}\varepsilon_1^2 + \varepsilon_2^2 + \ldots + \varepsilon_n^2 \quad \forall k = 0,1,\ldots,K \Rightarrow \begin{pmatrix} \frac{\partial}{\partial \beta_0}\vec{\varepsilon}^T\vec{\varepsilon} \\ \frac{\partial}{\partial \beta_1}\vec{\varepsilon}^T\vec{\varepsilon} \\ \ldots \\ \frac{\partial}{\partial \beta_K}\vec{\varepsilon}^T\vec{\varepsilon} \end{pmatrix}_{(K+1)\times 1} = \frac{\partial \vec{\varepsilon}^T\vec{\varepsilon}}{\partial \vec{\beta}} \overset{!}{=} \vec{0}$

Hence: $$\frac{\partial \vec{\varepsilon}^T\vec{\varepsilon}}{\partial \vec{\beta}} = \frac{\partial}{\partial \vec{\beta}}\left(\vec{y}^T\vec{y} - 2\vec{\beta}^T X^T \vec{y} + \vec{\beta}^T X^T X \vec{\beta}\right)$$

$$= \vec{0} - \underbrace{\frac{\partial}{\partial \vec{\beta}} 2\vec{\beta}^T X^T \vec{y}}_{2X^T\vec{y}} + \frac{\partial}{\partial \vec{\beta}}\vec{\beta}^T X^T X \vec{\beta}$$

$$\boxed{\frac{\partial}{\partial \beta_k}} 2\vec{\beta}^T_{1\times(K+1)} \underbrace{X^T\vec{y}}_{\vec{m}}{}_{(K+1)\times 1} = ??$$

Tutorial 7

PML SS 2025

Minimizing the Sum of Squared Errors

Sum of squared errors: $\vec{\varepsilon}^T\vec{\varepsilon} = \vec{y}^T\vec{y} - 2\vec{\beta}^T X^T \vec{y} + \vec{\beta}^T X^T X \vec{\beta} \in \mathbb{R}$

Derivatives/FOC: $\frac{\partial}{\partial \beta_k} \varepsilon_1^2 + \varepsilon_2^2 + \ldots + \varepsilon_n^2 \quad \forall k = 0,1,\ldots,K \Rightarrow \begin{pmatrix} \frac{\partial}{\partial \beta_0}\vec{\varepsilon}^T\vec{\varepsilon} \\ \frac{\partial}{\partial \beta_1}\vec{\varepsilon}^T\vec{\varepsilon} \\ \ldots \\ \frac{\partial}{\partial \beta_K}\vec{\varepsilon}^T\vec{\varepsilon} \end{pmatrix}_{(K+1)\times 1} = \frac{\partial \vec{\varepsilon}^T\vec{\varepsilon}}{\partial \vec{\beta}} \overset{!}{=} \vec{0}$

Hence:

$$\frac{\partial \vec{\varepsilon}^T\vec{\varepsilon}}{\partial \vec{\beta}} = \frac{\partial}{\partial \vec{\beta}}\left(\vec{y}^T\vec{y} - 2\vec{\beta}^T X^T \vec{y} + \vec{\beta}^T X^T X \vec{\beta}\right)$$

$$= \vec{0} - \underbrace{\frac{\partial}{\partial \vec{\beta}} 2\vec{\beta}^T X^T \vec{y}}_{2X^T\vec{y}} + \frac{\partial}{\partial \vec{\beta}} \vec{\beta}^T X^T X \vec{\beta}$$

$$\frac{\partial}{\partial \beta_k} 2\vec{\beta}^T_{1\times(K+1)} \underbrace{X^T\vec{y}}_{\vec{m}}{}_{(K+1)\times 1} = 2\frac{\partial}{\partial \beta_k}\left(\beta_0 \cdot m_0 \quad + \beta_1 \cdot m_1 \quad \ldots \quad + \beta_K \cdot m_K\right) = ??$$

Minimizing the Sum of Squared Errors

Sum of squared errors: $\vec{\varepsilon}^T\vec{\varepsilon} = \vec{y}^T\vec{y} - 2\vec{\beta}^T X^T \vec{y} + \vec{\beta}^T X^T X \vec{\beta} \in \mathbb{R}$

Derivatives/FOC:

$$\frac{\partial}{\partial \beta_k} \varepsilon_1^2 + \varepsilon_2^2 + \ldots + \varepsilon_n^2 \quad \forall k = 0,1,\ldots,K \quad \Rightarrow \begin{pmatrix} \frac{\partial}{\partial \beta_0}\vec{\varepsilon}^T\vec{\varepsilon} \\ \frac{\partial}{\partial \beta_1}\vec{\varepsilon}^T\vec{\varepsilon} \\ \ldots \\ \frac{\partial}{\partial \beta_K}\vec{\varepsilon}^T\vec{\varepsilon} \end{pmatrix}_{(K+1)\times 1} = \frac{\partial \vec{\varepsilon}^T\vec{\varepsilon}}{\partial \vec{\beta}} \overset{!}{=} \vec{0}$$

Hence:

$$\frac{\partial \vec{\varepsilon}^T\vec{\varepsilon}}{\partial \vec{\beta}} = \frac{\partial}{\partial \vec{\beta}}\left(\vec{y}^T\vec{y} - 2\vec{\beta}^T X^T \vec{y} + \vec{\beta}^T X^T X \vec{\beta}\right)$$

$$= \vec{0} - \underbrace{\frac{\partial}{\partial \vec{\beta}} 2\vec{\beta}^T X^T \vec{y}}_{2X^T\vec{y}} + \frac{\partial}{\partial \vec{\beta}} \vec{\beta}^T X^T X \vec{\beta}$$

$$\frac{\partial}{\partial \beta_k} 2\vec{\beta}^T_{1\times(K+1)} \underbrace{X^T\vec{y}}_{\vec{m}}{}_{(K+1)\times 1} = 2\frac{\partial}{\partial \beta_k}\left(\beta_0 \cdot m_0 \quad + \beta_1 \cdot m_1 \quad \ldots \quad + \beta_K \cdot m_K\right) = 2 \cdot m_k \qquad \forall k = 0,1,\ldots,K$$

Minimizing the Sum of Squared Errors

Hence: $$\frac{\partial \vec{\varepsilon}^T \vec{\varepsilon}}{\partial \vec{\beta}} = \frac{\partial}{\partial \vec{\beta}} \left(\vec{y}^T \vec{y} - 2\vec{\beta}^T X^T \vec{y} + \vec{\beta}^T X^T X \vec{\beta} \right)$$

$$= \vec{0} - 2X^T \vec{y} + \underbrace{\frac{\partial}{\partial \vec{\beta}} \vec{\beta}^T X^T X \vec{\beta}}_{(K+1)\times 1}$$

Minimizing the Sum of Squared Errors

$$\boxed{\frac{\partial}{\partial \beta_k}} \underbrace{\vec{\beta}^T}_{1\times(K+1)} \underbrace{\underbrace{X^T X}_{(K+1)\times(K+1)}}_{\begin{pmatrix} \vec{m}_0^T \\ \vec{m}_1^T \\ \cdots \\ \vec{m}_K^T \end{pmatrix}} \underbrace{\vec{\beta}}_{(K+1)\times 1} = \frac{\partial}{\partial \beta_k} \underbrace{\vec{\beta}^T}_{1\times(K+1)} \cdot \begin{pmatrix} \beta_0 \cdot m_{0,0} & +\beta_1 \cdot m_{0,1} & \ldots & +\beta_K \cdot m_{0,K} \\ \beta_0 \cdot m_{1,0} & +\beta_1 \cdot m_{1,1} & \ldots & +\beta_K \cdot m_{1,K} \\ & & \ldots & \\ \beta_0 \cdot m_{K,0} & +\beta_1 \cdot m_{K,1} & \ldots & +\beta_K \cdot m_{K,K} \end{pmatrix} = ??$$

Hence: $$\frac{\partial \vec{\varepsilon}^T \vec{\varepsilon}}{\partial \vec{\beta}} = \frac{\partial}{\partial \vec{\beta}} \left(\vec{y}^T \vec{y} - 2\vec{\beta}^T X^T \vec{y} + \vec{\beta}^T X^T X \vec{\beta} \right)$$

$$= \vec{0} - 2X^T \vec{y} + \underbrace{\frac{\partial}{\partial \vec{\beta}} \vec{\beta}^T X^T X \vec{\beta}}_{(K+1)\times 1}$$

Minimizing the Sum of Squared Errors

$$\frac{\partial}{\partial \beta_k} \underbrace{\vec{\beta}^T}_{1\times(K+1)} \underbrace{\underbrace{X^T X}_{(K+1)\times(K+1)}}_{\begin{pmatrix} \vec{m}_0^T \\ \vec{m}_1^T \\ \cdots \\ \vec{m}_K^T \end{pmatrix}} \underbrace{\vec{\beta}}_{(K+1)\times 1} = \frac{\partial}{\partial \beta_k} \underbrace{\vec{\beta}^T}_{1\times(K+1)} \cdot \begin{pmatrix} \beta_0 \cdot m_{0,0} & +\beta_1 \cdot m_{0,1} & \dots & +\beta_K \cdot m_{0,K} \\ \beta_0 \cdot m_{1,0} & +\beta_1 \cdot m_{1,1} & \dots & +\beta_K \cdot m_{1,K} \\ & & \dots & \\ \beta_0 \cdot m_{K,0} & +\beta_1 \cdot m_{K,1} & \dots & +\beta_K \cdot m_{K,K} \end{pmatrix} = \frac{\partial}{\partial \beta_k} \begin{pmatrix} \beta_0 \cdot \left(\beta_0 \cdot m_{0,0} \quad +\beta_1 \cdot m_{0,1} \quad \dots \quad +\beta_K \cdot m_{0,K}\right) \\ +\beta_1 \cdot \left(\beta_0 \cdot m_{1,0} \quad +\beta_1 \cdot m_{1,1} \quad \dots \quad +\beta_K \cdot m_{1,K}\right) \\ \dots \\ +\beta_K \cdot \left(\beta_0 \cdot m_{K,0} \quad +\beta_1 \cdot m_{K,1} \quad \dots \quad +\beta_K \cdot m_{K,K}\right) \end{pmatrix}$$

Hence: $$\frac{\partial \vec{\varepsilon}^T \vec{\varepsilon}}{\partial \vec{\beta}} = \frac{\partial}{\partial \vec{\beta}} \left(\vec{y}^T \vec{y} - 2 \vec{\beta}^T X^T \vec{y} + \vec{\beta}^T X^T X \vec{\beta} \right)$$

$$= \vec{0} - 2 X^T \vec{y} + \underbrace{\frac{\partial}{\partial \vec{\beta}} \vec{\beta}^T X^T X \vec{\beta}}_{(K+1)\times 1}$$

Tutorial 7

PML SS 2025

Minimizing the Sum of Squared Errors

$$\frac{\partial}{\partial \beta_k} \underbrace{\vec{\beta}^T}_{1\times(K+1)} \underbrace{\underbrace{X^T X}_{(K+1)\times(K+1)}}_{\begin{pmatrix} \vec{m}_0^T \\ \vec{m}_1^T \\ \dots \\ \vec{m}_K^T \end{pmatrix}} \underbrace{\vec{\beta}}_{(K+1)\times 1} = \frac{\partial}{\partial \beta_k} \underbrace{\vec{\beta}^T}_{1\times(K+1)} \cdot \begin{pmatrix} \beta_0 \cdot m_{0,0} & +\beta_1 \cdot m_{0,1} & \dots & +\beta_K \cdot m_{0,K} \\ \beta_0 \cdot m_{1,0} & +\beta_1 \cdot m_{1,1} & \dots & +\beta_K \cdot m_{1,K} \\ & & \dots & \\ \beta_0 \cdot m_{K,0} & +\beta_1 \cdot m_{K,1} & \dots & +\beta_K \cdot m_{K,K} \end{pmatrix} = \frac{\partial}{\partial \beta_k} \begin{pmatrix} \beta_0 \cdot \left(\beta_0 \cdot m_{0,0} \quad +\beta_1 \cdot m_{0,1} \quad \dots \quad +\beta_K \cdot m_{0,K}\right) \\ +\beta_1 \cdot \left(\beta_0 \cdot m_{1,0} \quad +\beta_1 \cdot m_{1,1} \quad \dots \quad +\beta_K \cdot m_{1,K}\right) \\ \dots \\ +\beta_K \cdot \left(\beta_0 \cdot m_{K,0} \quad +\beta_1 \cdot m_{K,1} \quad \dots \quad +\beta_K \cdot m_{K,K}\right) \end{pmatrix}$$

$$= \begin{pmatrix} \beta_0 \cdot m_{0,k} \\ +\beta_1 \cdot m_{1,k} \\ +1 \cdot \left(\beta_0 \cdot m_{k,0} \quad +\beta_1 \cdot m_{k,1} \quad \dots \quad +\beta_K \cdot m_{k,K}\right) + \beta_k \cdot m_{k,k} \\ +\beta_K \cdot m_{K,k} \end{pmatrix}$$

Hence: $$\frac{\partial \vec{\varepsilon}^T \vec{\varepsilon}}{\partial \vec{\beta}} = \frac{\partial}{\partial \vec{\beta}} \left(\vec{y}^T \vec{y} - 2\vec{\beta}^T X^T \vec{y} + \vec{\beta}^T X^T X \vec{\beta} \right)$$

$$= \vec{0} - 2X^T \vec{y} + \underbrace{\frac{\partial}{\partial \vec{\beta}} \vec{\beta}^T X^T X \vec{\beta}}_{(K+1)\times 1}$$

Minimizing the Sum of Squared Errors

$$\frac{\partial}{\partial \beta_k} \underbrace{\vec{\beta}^T}_{1\times(K+1)} \underbrace{\underbrace{X^T X}_{(K+1)\times(K+1)}}_{\begin{pmatrix} \vec{m}_0^T \\ \vec{m}_1^T \\ \dots \\ \vec{m}_K^T \end{pmatrix}} \underbrace{\vec{\beta}}_{(K+1)\times 1} = \frac{\partial}{\partial \beta_k} \underbrace{\vec{\beta}^T}_{1\times(K+1)} \cdot \begin{pmatrix} \beta_0 \cdot m_{0,0} & +\beta_1 \cdot m_{0,1} & \dots & +\beta_K \cdot m_{0,K} \\ \beta_0 \cdot m_{1,0} & +\beta_1 \cdot m_{1,1} & \dots & +\beta_K \cdot m_{1,K} \\ & & \dots & \\ \beta_0 \cdot m_{K,0} & +\beta_1 \cdot m_{K,1} & \dots & +\beta_K \cdot m_{K,K} \end{pmatrix} = \frac{\partial}{\partial \beta_k} \begin{pmatrix} \beta_0 \cdot \left(\beta_0 \cdot m_{0,0} \quad +\beta_1 \cdot m_{0,1} \quad \dots \quad +\beta_K \cdot m_{0,K}\right) \\ +\beta_1 \cdot \left(\beta_0 \cdot m_{1,0} \quad +\beta_1 \cdot m_{1,1} \quad \dots \quad +\beta_K \cdot m_{1,K}\right) \\ \dots \\ +\beta_K \cdot \left(\beta_0 \cdot m_{K,0} \quad +\beta_1 \cdot m_{K,1} \quad \dots \quad +\beta_K \cdot m_{K,K}\right) \end{pmatrix}$$

$$= \begin{pmatrix} \beta_0 \cdot m_{0,k} \\ +\beta_1 \cdot m_{1,k} \\ +1 \cdot \left(\beta_0 \cdot m_{k,0} \quad +\beta_1 \cdot m_{k,1} \quad \dots \quad +\beta_K \cdot m_{k,K}\right) + \beta_k \cdot m_{k,k} \\ +\beta_K \cdot m_{K,k} \end{pmatrix}$$

Hence: $$\frac{\partial \vec{\varepsilon}^T \vec{\varepsilon}}{\partial \vec{\beta}} = \frac{\partial}{\partial \vec{\beta}} \left(\vec{y}^T \vec{y} - 2\vec{\beta}^T X^T \vec{y} + \vec{\beta}^T X^T X \vec{\beta} \right)$$

$$= \vec{0} - 2X^T \vec{y} + \underbrace{\frac{\partial}{\partial \vec{\beta}} \vec{\beta}^T X^T X \vec{\beta}}_{\underbrace{2X^T X \vec{\beta}}_{(K+1)\times 1}}$$

$$\boxed{\begin{array}{l} X^T X \text{ symmetric} \\ m_{k,i} = m_{i,k}, \forall i,k = 0,1,\dots,K \end{array}}$$

$$= 2 \cdot \left(\beta_0 \cdot m_{k,0} \quad +\beta_1 \cdot m_{k,1} \quad \dots \quad +\beta_K \cdot m_{k,K}\right)$$

$$= 2 \cdot k^{th} \text{ row of } X^T X \vec{\beta} \qquad \forall k = 0,1,\dots,K$$

Minimizing the Sum of Squared Errors

Sum of squared errors: $\vec{\varepsilon}^T\vec{\varepsilon} = \vec{y}^T\vec{y} - 2\vec{\beta}^T X^T\vec{y} + \vec{\beta}^T X^T X\vec{\beta} \in \mathbb{R}$

Derivatives/FOC: $\frac{\partial}{\partial \beta_k}\varepsilon_1^2 + \varepsilon_2^2 + \ldots + \varepsilon_n^2 \quad \forall k = 0,1,\ldots,K \Rightarrow \begin{pmatrix} \frac{\partial}{\partial \beta_0}\vec{\varepsilon}^T\vec{\varepsilon} \\ \frac{\partial}{\partial \beta_1}\vec{\varepsilon}^T\vec{\varepsilon} \\ \ldots \\ \frac{\partial}{\partial \beta_K}\vec{\varepsilon}^T\vec{\varepsilon} \end{pmatrix}_{(K+1)\times 1} = \frac{\partial \vec{\varepsilon}^T\vec{\varepsilon}}{\partial \vec{\beta}} \overset{!}{=} \vec{0}$

Hence:

$$\frac{\partial \vec{\varepsilon}^T\vec{\varepsilon}}{\partial \vec{\beta}} = \frac{\partial}{\partial \vec{\beta}}\left(\vec{y}^T\vec{y} - 2\vec{\beta}^T X^T\vec{y} + \vec{\beta}^T X^T X\vec{\beta}\right)$$

$$= \frac{\partial}{\partial \vec{\beta}}\vec{y}^T\vec{y} - \frac{\partial}{\partial \vec{\beta}}2\vec{\beta}^T X^T\vec{y} + \frac{\partial}{\partial \vec{\beta}}\vec{\beta}^T X^T X\vec{\beta}$$

$$= \boxed{-2X^T\vec{y} + 2X^T X\vec{\beta} \overset{!}{=} \vec{0}}$$

Minimizing the Sum of Squared Errors

Sum of squared errors: $\vec{\varepsilon}^T\vec{\varepsilon} = \vec{y}^T\vec{y} - 2\vec{\beta}^T X^T \vec{y} + \vec{\beta}^T X^T X \vec{\beta} \in \mathbb{R}$

Derivatives/FOC: $\frac{\partial}{\partial \beta_k} \varepsilon_1^2 + \varepsilon_2^2 + \ldots + \varepsilon_n^2 \quad \forall k = 0,1,\ldots,K \quad \Rightarrow \begin{pmatrix} \frac{\partial}{\partial \beta_0}\vec{\varepsilon}^T\vec{\varepsilon} \\ \frac{\partial}{\partial \beta_1}\vec{\varepsilon}^T\vec{\varepsilon} \\ \ldots \\ \frac{\partial}{\partial \beta_K}\vec{\varepsilon}^T\vec{\varepsilon} \end{pmatrix}_{(K+1)\times 1} = \frac{\partial \vec{\varepsilon}^T\vec{\varepsilon}}{\partial \vec{\beta}} \overset{!}{=} \vec{0}$

Hence:

$$\frac{\partial \vec{\varepsilon}^T\vec{\varepsilon}}{\partial \vec{\beta}} = \frac{\partial}{\partial \vec{\beta}}\left(\vec{y}^T\vec{y} - 2\vec{\beta}^T X^T \vec{y} + \vec{\beta}^T X^T X \vec{\beta}\right)$$

$$= \frac{\partial}{\partial \vec{\beta}}\vec{y}^T\vec{y} - \frac{\partial}{\partial \vec{\beta}} 2\vec{\beta}^T X^T \vec{y} + \frac{\partial}{\partial \vec{\beta}}\vec{\beta}^T X^T X \vec{\beta}$$

$$= -2X^T\vec{y} + 2X^T X\vec{\beta} \overset{!}{=} \vec{0}$$

$$\Leftrightarrow \quad X^T X \vec{\beta} \overset{!}{=} X^T \vec{y}$$

What do we want again?

Minimizing the Sum of Squared Errors

Optimality condition: $$\underbrace{X^T}_{(K+1)\times\ldots} X \underbrace{\vec{\beta}}_{\ldots\times 1} \stackrel{!}{=} \underbrace{X^T \vec{y}}_{(K+1)\times 1}$$

How to get the optimal weights $\vec{\beta}$?

Any ideas?

Minimizing the Sum of Squared Errors

Optimality condition:
$$\underbrace{X^T}_{(K+1)\times\ldots} X \underbrace{\vec{\beta}}_{\ldots\times 1} \overset{!}{=} \underbrace{X^T \vec{y}}_{(K+1)\times 1}$$

How to get the optimal weights $\vec{\beta}$?

If $X^T X$ has full rank, then the inverse exists and we get:

$$\Leftrightarrow \quad \left(X^T X\right)^{-1} \cdot X^T X \vec{\beta} = \left(X^T X\right)^{-1} \cdot X^T \vec{y}$$

Minimizing the Sum of Squared Errors

Optimality condition:
$$\underbrace{X^T}_{(K+1)\times\ldots} X \underbrace{\vec{\beta}}_{\ldots\times 1} \overset{!}{=} \underbrace{X^T \vec{y}}_{(K+1)\times 1}$$

How to get the optimal weights $\vec{\beta}$?

If $X^T X$ has full rank, then the inverse exists and we get:

$$\Leftrightarrow \quad \left(X^T X\right)^{-1} \cdot X^T X \vec{\beta} = \left(X^T X\right)^{-1} \cdot X^T \vec{y}$$

$$\Leftrightarrow \quad I \cdot \vec{\beta} = \left(X^T X\right)^{-1} \cdot X^T \vec{y}$$

$$\Leftrightarrow \quad \boxed{\vec{\beta} = \left(X^T X\right)^{-1} \cdot X^T \vec{y}}$$

Minimizing the Sum of Squared Errors

Optimality condition: $$\underbrace{X^T}_{(K+1)\times\ldots} X \underbrace{\vec{\beta}}_{\ldots\times 1} \overset{!}{=} \underbrace{X^T \vec{y}}_{(K+1)\times 1}$$ ← Other ideas to get $\vec{\beta}$??

How to get the optimal weights $\vec{\beta}$?

If $X^T X$ has full rank, then the inverse exists and we get:

$$\Leftrightarrow \quad \left(X^T X\right)^{-1} \cdot X^T X \vec{\beta} = \left(X^T X\right)^{-1} \cdot X^T \vec{y}$$

$$\Leftrightarrow \quad I \cdot \vec{\beta} = \left(X^T X\right)^{-1} \cdot X^T \vec{y}$$

$$\Leftrightarrow \quad \vec{\beta} = \left(X^T X\right)^{-1} \cdot X^T \vec{y}$$

Minimizing the Sum of Squared Errors

Optimality condition:

$$\underbrace{X^T}_{(K+1)\times\ldots} X \underbrace{\vec{\beta}}_{\ldots\times 1} \stackrel{!}{=} \underbrace{X^T \vec{y}}_{(K+1)\times 1} \longleftarrow$$

or solve $A\vec{x} = \vec{b}$ *for* $\vec{x}$

where $A = X^T X$

and $\vec{x} = \vec{\beta},\ \vec{b} = X^T \vec{y}$

How to get the optimal weights $\vec{\beta}$?

If $X^T X$ has full rank, then the inverse exists and we get:

$$\Leftrightarrow \quad \left(X^T X\right)^{-1} \cdot X^T X \vec{\beta} = \left(X^T X\right)^{-1} \cdot X^T \vec{y}$$

$$\Leftrightarrow \quad I \cdot \vec{\beta} = \left(X^T X\right)^{-1} \cdot X^T \vec{y}$$

$$\Leftrightarrow \quad \vec{\beta} = \left(X^T X\right)^{-1} \cdot X^T \vec{y}$$

Summary

- Recap I: Cholesky Decomposition
- Recap II: Linear Basis Function Models
- Recap III: Linear Regression & Matrix Algebra

$\mathbf{y}$	$\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$
$\mathbf{A}\mathbf{x}$	$\mathbf{A}^T$
$\mathbf{x}^T\mathbf{A}$	$\mathbf{A}$
$\mathbf{x}^T\mathbf{x}$	$2\mathbf{x}$
$\mathbf{x}^T\mathbf{A}\mathbf{x}$	$\mathbf{A}\mathbf{x} + \mathbf{A}^T\mathbf{x}$

$$
\begin{aligned}
(A + B)^T &= A^T + B^T \\
(c \cdot A)^T &= c \cdot A^T \\
(A^T)^T &= A \\
(A \cdot B)^T &= B^T \cdot A^T \\
(A^{-1})^T &= (A^T)^{-1}
\end{aligned}
$$

See you next Week!