Simple Linear Regression

Matrix Representation

Ravleen Bajaj

Today:

- Verify the six facts discussed in Lecture 14, matrices, derivatives, and expectations by writing out scalar-level calculations.
- Characterization and marginalization of the MVN distribution

Define

We will use these vectors throughout the verifications:

$$\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}, \quad \mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \end{pmatrix}_{2\times 3}.$$

Fact 1:

$$\frac{\partial}{\partial \mathbf{w}} \mathbf{w}' \ \mathbf{z} = \mathbf{z}.$$

Fact 1:

$$\frac{\partial}{\partial \mathbf{w}} \mathbf{w'} \ \mathbf{z} = \mathbf{z}.$$

Verification:

$$\mathbf{w'z} = \begin{pmatrix} w_1 & w_2 & w_3 \end{pmatrix} \cdot \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = w_1 z_1 + w_2 z_2 + w_3 z_3 = \sum_{i=1}^{3} w_i z_i$$

$$\frac{\partial}{\partial w_i}(w_1 z_1 + w_2 z_2 + w_3 z_3) = z_i$$

Generally,

$$\mathbf{w'z} = \sum_{i=1}^{n} w_i z_i$$

Fact 2:

$$\mathbf{w}'\ \mathbf{z} = \mathbf{z}'\ \mathbf{w}.$$

Fact 2:

$$\mathbf{w'}\ \mathbf{z} = \mathbf{z'}\ \mathbf{w}.$$

Verification: Both sides are

$$\mathbf{w'}\ \mathbf{z} = \begin{pmatrix} w_1 & w_2 & w_3 \end{pmatrix} \cdot \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = w_1 z_1 + w_2 z_2 + w_3 z_3 = \sum_{i=1}^{3} w_i z_i,$$

and

$$\mathbf{z'}\ \mathbf{w} = \begin{pmatrix} z_1 & z_2 & z_3 \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = z_1 w_1 + z_2 w_2 + z_3 w_3 = \sum_{i=1}^{3} w_i z_i.$$

Generally, $\mathbf{w'z} = \sum_{i=1}^{n} w_i z_i = \sum_{i=1}^{n} z_i w_i = \mathbf{z'w}$

Fact 3:

$$\frac{\partial}{\partial \mathbf{w}}\left(\mathbf{w'A'Aw}\right) = 2\mathbf{A'}\ \mathbf{A}\ \mathbf{w}.$$

Fact 3:

$$\frac{\partial}{\partial \mathbf{w}}\left(\mathbf{w'A'Aw}\right) = 2\mathbf{A'}\ \mathbf{A}\ \mathbf{w}.$$

Step 1: Compute $\mathbf{w'(A'A)w}$:

$$\mathbf{w'(A'A)w} = \mathbf{w'}\begin{pmatrix}1 & 0\\ 0 & 1\\ 2 & 1\end{pmatrix}\begin{pmatrix}1 & 0 & 2\\ 0 & 1 & 1\end{pmatrix}\mathbf{w}$$

$$= \begin{pmatrix}w_1 & w_2 & w_3\end{pmatrix}\begin{pmatrix}1 & 0 & 2\\ 0 & 1 & 1\\ 2 & 1 & 5\end{pmatrix}\begin{pmatrix}w_1\\ w_2\\ w_3\end{pmatrix}$$

$$= w_1(1w_1 + 0w_2 + 2w_3) + w_2(0w_1 + 1w_2 + 1w_3) +$$
$$w_3(2w_1 + 1w_2 + 5w_3)$$

$$= w_1^2 + w_2^2 + 5w_3^2 + 4w_1w_3 + 2w_2w_3.$$

Fact 3 (Contd.):

Step 2: Compute 2**A' A w**:

$$2\mathbf{A'Aw} = 2\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 2 & 1 & 5 \end{pmatrix}\begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$$

$$= \begin{pmatrix} 2 & 0 & 4 \\ 0 & 2 & 2 \\ 4 & 2 & 10 \end{pmatrix}\begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}$$

$$= \begin{pmatrix} 2w_1 + 4w_3 \\ 2w_2 + 2w_3 \\ 4w_1 + 2w_2 + 10w_3 \end{pmatrix}$$

Fact 3 (Contd.):

Step 3: Compute $\frac{\partial}{\partial w_i}\left(w_1^2 + w_2^2 + 5w_3^2 + 4w_1w_3 + 2w_2w_3\right)$

1. $\frac{\partial}{\partial w_1}\left(w_1^2 + w_2^2 + 5w_3^2 + 4w_1w_3 + 2w_2w_3\right) = 2w_1 + 4w_3$
2. $\frac{\partial}{\partial w_2}\left(w_1^2 + w_2^2 + 5w_3^2 + 4w_1w_3 + 2w_2w_3\right) = 2w_2 + 2w_3$
3. $\frac{\partial}{\partial w_3}\left(w_1^2 + w_2^2 + 5w_3^2 + 4w_1w_3 + 2w_2w_3\right) = 4w_1 + 2w_2 + 10w_3$

$$\Rightarrow \quad \frac{\partial}{\partial \mathbf{w}}\left(\mathbf{w'A'Aw}\right) = \begin{pmatrix} 2w_1 + 4w_3 \\ 2w_2 + 2w_3 \\ 4w_1 + 2w_2 + 10w_3 \end{pmatrix}$$

Fact 3: Generalization

Let

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{np} \end{pmatrix}_{n \times p}, \quad \mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{pmatrix}.$$

Then

$$\mathbf{A}\mathbf{w} = \begin{pmatrix} a_{11}w_1 + a_{12}w_2 + \dots + a_{1p}w_p \\ a_{21}w_1 + a_{22}w_2 + \dots + a_{2p}w_p \\ \vdots \\ a_{n1}w_1 + a_{n2}w_2 + \dots + a_{np}w_p \end{pmatrix}.$$

Fact 3: Generalization (Contd.)

$$\mathbf{w}'(A'A)\mathbf{w} = (A\mathbf{w})'(A\mathbf{w}).$$

That is, **w'(A'A)w** is the *inner product of* **Aw** *with itself*.

$$\boxed{\mathbf{w'(A'A)}'\mathbf{w} = (\mathbf{Aw})'(\mathbf{Aw}) = \|\mathbf{Aw}\|^2 = \sum_{i=1}^{n}(a_{i1}w_1 + a_{i2}w_2 + \cdots + a_{ip}w_p)^2.}$$

Fact 3: Generalization (Contd.)

Expanding the sum gives:

$$\mathbf{w}'(A'A)\mathbf{w} = \sum_{i=1}^{n}\sum_{j=1}^{p}\sum_{k=1}^{p} a_{ij}a_{ik}w_j w_k.$$

Grouping by j, k:

$$\boxed{\mathbf{w}'(A'A)\mathbf{w} = \sum_{j=1}^{p}\sum_{k=1}^{p}\left(\sum_{i=1}^{n} a_{ij}a_{ik}\right) w_j w_k.}$$

The coefficient of $w_j w_k$ is the $(j,k)^{th}$ entry of $A'A$:

$$(A'A)_{jk} = \sum_{i=1}^{n} a_{ij}a_{ik}.$$

Fact 4:

For a random column vector $\mathbf{Z} = (Z_1, Z_2)'$ and constant matrix $\mathbf{A}$,

$$E[\mathbf{Z}] = \big(E[Z_1], E[Z_2]\big)', \qquad E[\mathbf{AZ}] = \mathbf{A}\, E[\mathbf{Z}].$$

Verification: Compute the expectation componentwise:

$$E[\mathbf{Z}] = \big(E[Z_1], \quad E[Z_2]\big)\,.$$

Fact 4 (Contd.):

Let A_i be the i^{th} row of A. Then we need to compute $E[\mathbf{A}_i\mathbf{Z}]$ for each i: So

$$E[\mathbf{AZ}] = \begin{pmatrix} E[A_1 Z_1] \\ E[A_2 Z_2] \end{pmatrix},$$

while

$$\mathbf{A}\, E[Z] = \mathbf{A} E[\mathbf{Z}] = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \left(E[Z_1], \quad E[Z_2] \right) = \begin{pmatrix} A_1 E[Z_1] \\ A_2 E[Z_2] \end{pmatrix}.$$

Hence $E[\mathbf{AZ}] = \mathbf{A}\, E[\mathbf{Z}]$.

Fact 4: Generalization

Let

$$\mathbf{Z} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{pmatrix}, \qquad A = (a_{ij})_{n\times p}.$$

Then each component of **AZ** is

$$(\mathbf{AZ})_i = \sum_{j=1}^{p} a_{ij}\, Z_j.$$

Fact 4: Generalization (Contd.)

By the linearity of expectation,

$$E[(\mathbf{AZ})_i] = \sum_{j=1}^{p} a_{ij}\, E[Z_j].$$

Stacking these

$$E[\mathbf{AZ}] = \mathbf{A} \begin{pmatrix} E[Z_1] \\ E[Z_2] \\ \vdots \\ E[Z_p] \end{pmatrix} = \mathbf{A}\, E[\mathbf{Z}].$$

Fact 5:

A'A is symmetric. If **A'A** is invertible, its inverse is symmetric.

Fact 5:

A'A is symmetric. If **A'A** is invertible, its inverse is symmetric.

We know,

$$\mathbf{A'A} = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 2 & 1 & 5 \end{pmatrix}.$$

Check symmetry by transposing:

$$\mathbf{A'A}' = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 2 & 1 & 5 \end{pmatrix}' = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 2 & 1 & 5 \end{pmatrix} = \mathbf{A'A}.$$

So $A'A$ is symmetric.

Here $\det(\mathbf{A'A}) = 0$, so **A'A** is singular and *does not* have an inverse. Intuition: **A** is 2×3 so $\text{rank}(\mathbf{A}) \leq 2 < 3$, hence **A'A** (a 3×3 matrix) cannot be full rank and is singular.

Fact 5: Generalization

Recall that for any matrices B and C with compatible dimensions,

$$(BC)' = C'B'.$$

Apply this property with $B = \mathbf{A}'$ and $C = \mathbf{A}$.
Then

$$(\mathbf{A}'\mathbf{A})' = \mathbf{A}'(\mathbf{A}')' = \mathbf{A}'\mathbf{A}.$$

Thus $\mathbf{A}'\mathbf{A}$ is symmetric.

Fact 5: Generalization (Contd.)

If A′A is invertible:
Let $\mathbf{B} = \mathbf{A}'\mathbf{A}$. Then,

$$(\mathbf{B}^{-1})' = (\mathbf{B}')^{-1} = \mathbf{B}^{-1}.$$

Hence, the inverse of a symmetric and invertible matrix is also symmetric.

Fact 6: Generalization

$$Var[\mathbf{AZ}] = \mathbf{A}\, Var[\mathbf{Z}]\, \mathbf{A}'.$$

By definition:

$$Var[\mathbf{Z}] = Cov[\mathbf{Z}] = \begin{pmatrix} Var[Z_1] & Cov[Z_1, Z_2] & \cdots & Cov[Z_1, Z_p] \\ Cov[Z_2, Z_1] & Var[Z_2] & \cdots & Cov[Z_2, Z_p] \\ \vdots & \vdots & \ddots & \vdots \\ Cov[Z_p, Z_1] & Cov[Z_p, Z_2] & \cdots & Var[Z_p] \end{pmatrix}.$$

Fact 6: Generalization (contd.)

By definition of variance for a random vector,

$$Var[\mathbf{AZ}] = E\left[(\mathbf{AZ} - E[\mathbf{AZ}])(\mathbf{AZ} - E[\mathbf{AZ}])'\right].$$

Since $\mathbf{A}$ is constant and $E[\mathbf{AZ}] = \mathbf{A}E[\mathbf{Z}]$,

$$Var[\mathbf{AZ}] = E\left[\mathbf{A}(\mathbf{Z} - E[\mathbf{Z}])(\mathbf{Z} - E[\mathbf{Z}])'\mathbf{A}'\right].$$

Using the linearity of expectation and the fact that $\mathbf{A}$ is non-random,

$$Var[\mathbf{AZ}] = \mathbf{A}\, E\left[(\mathbf{Z} - E[\mathbf{Z}])(\mathbf{Z} - E[\mathbf{Z}])'\right] \mathbf{A}'.$$

But

$$E\left[(\mathbf{Z} - E[\mathbf{Z}])(\mathbf{Z} - E[\mathbf{Z}])'\right] = Var[\mathbf{Z}].$$

Hence,

$$Var[\mathbf{AZ}] = \mathbf{A}\, Var[\mathbf{Z}]\, \mathbf{A}'.$$

Fact 6:

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \end{pmatrix}, \qquad \mathbf{Z} = \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \end{pmatrix}.$$

Suppose

$$Var[\mathbf{Z}] = \begin{pmatrix} 1 & 0.2 & 0.4 \\ 0.2 & 1 & 0.3 \\ 0.4 & 0.3 & 2 \end{pmatrix}.$$

Compute

$$Var[\mathbf{AZ}] = \mathbf{A}\, Var[\mathbf{Z}]\, \mathbf{A}'.$$

Fact 6 (contd.):

First multiply:

$$\mathbf{A}\, Var[\mathbf{Z}] = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0.2 & 0.4 \\ 0.2 & 1 & 0.3 \\ 0.4 & 0.3 & 2 \end{pmatrix} = \begin{pmatrix} 1.8 & 0.8 & 4.4 \\ 0.6 & 1.3 & 2.3 \end{pmatrix}.$$

Then multiply by $\mathbf{A}'$:

$$\mathbf{A}\, Var[\mathbf{Z}]\mathbf{A}' = \begin{pmatrix} 1.8 & 0.8 & 4.4 \\ 0.6 & 1.3 & 2.3 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 10.6 & 5.2 \\ 5.2 & 3.6 \end{pmatrix}.$$

Hence

$$\mathbf{A}\, Var[\mathbf{Z}]\mathbf{A}' = \begin{pmatrix} 10.6 & 5.2 \\ 5.2 & 3.6 \end{pmatrix},$$

Fact 6 (contd.):

Let the rows of $\mathbf{A}$ be

$$\mathbf{a_1} = (1, 0, 2), \qquad \mathbf{a_2} = (0, 1, 1).$$

Then

$$Var[\mathbf{AZ}] = \begin{pmatrix} Var[\mathbf{a_1 Z}] & Cov[\mathbf{a_1 Z}, \mathbf{a_2 Z}] \\ Cov[\mathbf{a_2 Z}, \mathbf{a_1 Z}] & Var[\mathbf{a_2 Z}] \end{pmatrix}.$$

Compute each term:

$$Var[\mathbf{a_1 Z}] = \mathbf{a_1}\, Var[\mathbf{Z}]\, \mathbf{a}_1' = (1, 0, 2) \begin{pmatrix} 1 & 0.2 & 0.4 \\ 0.2 & 1 & 0.3 \\ 0.4 & 0.3 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} = 10.6,$$

$$Var[\mathbf{a_2 Z}] = (0, 1, 1)\, Var[\mathbf{Z}]\, (0, 1, 1)' = 3.6,$$

$$Cov[\mathbf{a_1 Z}, \mathbf{a_2 Z}] = (1, 0, 2)\, Var[\mathbf{Z}]\, (0, 1, 1)' = 5.2.$$

Multivariate Normality of $\mathbf{Y}$

- $\mathbf{Y}$ is a linear function of normally distributed random variables and is therefore **multivariate normal distributed**

$$\mathbf{Y} = X\boldsymbol{\beta} + \varepsilon, \quad \varepsilon \sim N(\mathbf{0}, \boldsymbol{\Lambda})$$

$$\boldsymbol{\Lambda} = \begin{pmatrix} \sigma^2 & 0 & \cdots \\ 0 & \sigma^2 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

- The variance-covariance matrix of $\mathbf{Y}$ is

$$Var[\mathbf{Y}] = Var[\varepsilon] = \boldsymbol{\Lambda}$$

- The expected value of $\mathbf{Y}$ is

$$E[\mathbf{Y}] = X\boldsymbol{\beta} + E[\varepsilon] = X\boldsymbol{\beta}$$

- Hence, $\mathbf{Y} \sim N(X\boldsymbol{\beta}, \boldsymbol{\Lambda})$

Multivariate Normal Distribution

A random vector

$$\mathbf{Z} = \begin{pmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_p \end{pmatrix}$$

is said to have a **multivariate normal distribution** with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ if its density is

$$f(\mathbf{z}) = \frac{1}{(2\pi)^{p/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\tfrac{1}{2}(\mathbf{z}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{z}-\boldsymbol{\mu})\right\}.$$

We write

$$\mathbf{Z} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

Characterization by Mean and Covariance

The distribution of a multivariate normal vector $\mathbf{Z}$ is **completely determined** by:

$$\boldsymbol{\mu} = \begin{pmatrix} E[Z_1] \\ E[Z_2] \\ \vdots \\ E[Z_p] \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} Var(Z_1) & Cov(Z_1, Z_2) & \cdots \\ Cov(Z_2, Z_1) & Var(Z_2) & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}.$$

- $\boldsymbol{\mu}$ gives the center of the distribution.
- $\boldsymbol{\Sigma}$ gives the spread of the distribution.
- Every linear combination of $\mathbf{Z}$ with a scalar is normally distributed.

Independence and Factorization

- If $Z_1, Z_2, \ldots, Z_p$ are **independent**, then $\boldsymbol{\Sigma}$ is diagonal and vice-versa only in case of normal density.

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_p^2 \end{pmatrix}$$

- The joint density factors as

$$f(\mathbf{z}) = \prod_{i=1}^{p} \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left\{ -\frac{1}{2} \left(\frac{z_i - \mu_i}{\sigma_i} \right)^2 \right\}$$

- Each Z_i is an independent $N(\mu_i, \sigma_i^2)$

Thank You!

Questions?