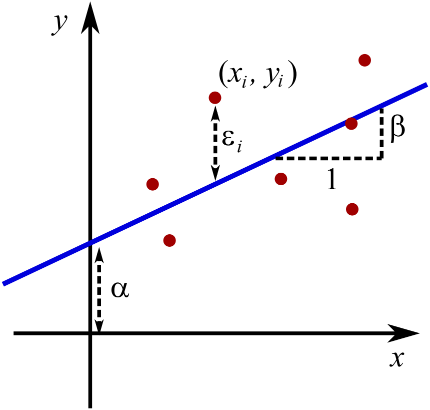
Linear Algebra Basics: Matrices

Foundations of Data Analysis

February 25, 2020

Review: Linear Regression

Model the data as a line:



$$y_i = \alpha + \beta x_i + \epsilon_i$$

α : intercept

β : slope

ϵ_i : error

Review: Least Squares

Goal: minimize **sum-of-squared error**:

$$\mathrm{SSE}(\alpha, \beta) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$

Review: Least Squares

Goal: minimize **sum-of-squared error**:

$$\text{SSE}(\alpha, \beta) = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$

Solution:

$$\hat{\beta} = \frac{\langle \tilde{x}, \tilde{y} \rangle}{\|\tilde{x}\|^2}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

Reminder: Centering Data

Means:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

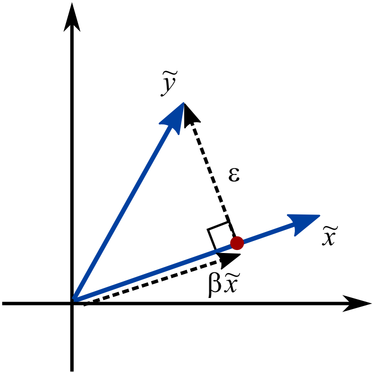
Centered data:

$$\tilde{y}_i = y_i - \bar{y}$$

$$\tilde{x}_i = x_i - \bar{x}$$

Review: Linear Regression as Projection

$$\begin{bmatrix} \tilde{y}_1 \\ \tilde{y}_2 \\ \vdots \\ \tilde{y}_n \end{bmatrix} = \beta \begin{bmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \vdots \\ \tilde{x}_n \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$



$$\hat{\beta} = \frac{\langle \tilde{x}, \tilde{y} \rangle}{\|\tilde{x}\|^2}$$

Review: Correlation vs. Regression Slope

Correlation:

$$\mathrm{corr}(x, y) = \frac{\langle \tilde{x}, \tilde{y} \rangle}{\|\tilde{x}\| \, \|\tilde{y}\|} = \cos\theta$$

Regression Slope:

$$\hat{\beta} = \frac{\langle \tilde{x}, \tilde{y} \rangle}{\|\tilde{x}\|^2} = \mathrm{corr}(x, y) \frac{\|\tilde{y}\|}{\|\tilde{x}\|}$$

R^2 Statistic

R^2 statistic gives the **proportion of explained variance**:

$$
\begin{aligned}
R^2 &= \frac{\text{explained variance in } y}{\text{total variance of } y} \\
&= 1 - \frac{\text{unexplained variance in } y}{\text{total variance of } y}
\end{aligned}
$$

R^2 Statistic

R^2 statistic gives the **proportion of explained variance**:

$$R^2 = \frac{\text{explained variance in } y}{\text{total variance of } y} = 1 - \frac{\text{unexplained variance in } y}{\text{total variance of } y}$$

$$R^2 = 1 - \frac{\mathrm{var}(\epsilon)}{\mathrm{var}(y)} = 1 - \frac{||\epsilon||^2}{||\tilde{y}||^2}$$

R^2 Statistic

R^2 statistic gives the **proportion of explained variance**:

$$R^2 = \frac{\text{explained variance in } y}{\text{total variance of } y}$$
$$= 1 - \frac{\text{unexplained variance in } y}{\text{total variance of } y}$$

$$R^2 = 1 - \frac{\mathrm{var}(\epsilon)}{\mathrm{var}(y)} = 1 - \frac{||\epsilon||^2}{||\tilde{y}||^2}$$

Range: R^2 is always between 0 and 1

Data Tables

	ID	M.F	Hand	Age	Educ	SES	MMSE	CDR	eTIV	nWBV	ASF	Delay	RightHippoVol	LeftHippoVol
0	OAS1_0002_MR1	F	R	55	4	1.0	29	0.0	1147	0.810	1.531	NaN	4230	3807
1	OAS1_0003_MR1	F	R	73	4	3.0	27	0.5	1454	0.708	1.207	NaN	2896	2801
2	OAS1_0010_MR1	M	R	74	5	2.0	30	0.0	1636	0.689	1.073	NaN	2832	2578
3	OAS1_0011_MR1	F	R	52	3	2.0	30	0.0	1321	0.827	1.329	NaN	3978	4080
4	OAS1_0013_MR1	F	R	81	5	2.0	30	0.0	1664	0.679	1.055	NaN	3557	3495
5	OAS1_0015_MR1	M	R	76	2	NaN	28	0.5	1738	0.719	1.010	NaN	3052	2770
6	OAS1_0016_MR1	M	R	82	2	4.0	27	0.5	1477	0.739	1.188	NaN	3421	3119
7	OAS1_0018_MR1	M	R	39	3	4.0	28	0.0	1636	0.813	1.073	NaN	4496	4283
8	OAS1_0019_MR1	F	R	89	5	1.0	30	0.0	1536	0.715	1.142	NaN	3760	3167
9	OAS1_0020_MR1	F	R	48	5	2.0	29	0.0	1326	0.785	1.323	NaN	3557	3394
10	OAS1_0021_MR1	F	R	80	3	3.0	23	0.5	1794	0.765	0.978	NaN	3715	3019
11	OAS1_0022_MR1	F	R	69	2	4.0	23	0.5	1447	0.757	1.213	NaN	3258	3566
12	OAS1_0023_MR1	M	R	82	2	3.0	27	0.5	1420	0.710	1.236	NaN	3217	2160
13	OAS1_0026_MR1	F	R	58	5	1.0	30	0.0	1235	0.820	1.421	NaN	3783	3535
14	OAS1_0028_MR1	F	R	86	2	4.0	27	1.0	1449	0.738	1.211	NaN	3452	3100
15	OAS1_0030_MR1	F	R	65	2	3.0	29	0.0	1392	0.764	1.261	NaN	3969	3406

Data Tables

	ID	M.F	Hand	Age	Educ	SES	MMSE	CDR	eTIV	nWBV	ASF	Delay	RightHippoVol	LeftHippoVol
0	OAS1_0002_MR1	F	R	55	4	1.0	29	0.0	1147	0.810	1.531	NaN	4230	3807
1	OAS1_0003_MR1	F	R	73	4	3.0	27	0.5	1454	0.708	1.207	NaN	2896	2801
2	OAS1_0010_MR1	M	R	74	5	2.0	30	0.0	1636	0.689	1.073	NaN	2832	2578
3	OAS1_0011_MR1	F	R	52	3	2.0	30	0.0	1321	0.827	1.329	NaN	3978	4080
4	OAS1_0013_MR1	F	R	81	5	2.0	30	0.0	1664	0.679	1.055	NaN	3557	3495
5	OAS1_0015_MR1	M	R	76	2	NaN	28	0.5	1738	0.719	1.010	NaN	3052	2770
6	OAS1_0016_MR1	M	R	82	2	4.0	27	0.5	1477	0.739	1.188	NaN	3421	3119
7	OAS1_0018_MR1	M	R	39	3	4.0	28	0.0	1636	0.813	1.073	NaN	4496	4283
8	OAS1_0019_MR1	F	R	89	5	1.0	30	0.0	1536	0.715	1.142	NaN	3760	3167
9	OAS1_0020_MR1	F	R	48	5	2.0	29	0.0	1326	0.785	1.323	NaN	3557	3394
10	OAS1_0021_MR1	F	R	80	3	3.0	23	0.5	1794	0.765	0.978	NaN	3715	3019
11	OAS1_0022_MR1	F	R	69	2	4.0	23	0.5	1447	0.757	1.213	NaN	3258	3566
12	OAS1_0023_MR1	M	R	82	2	3.0	27	0.5	1420	0.710	1.236	NaN	3217	2160
13	OAS1_0026_MR1	F	R	58	5	1.0	30	0.0	1235	0.820	1.421	NaN	3783	3535
14	OAS1_0028_MR1	F	R	86	2	4.0	27	1.0	1449	0.738	1.211	NaN	3452	3100
15	OAS1_0030_MR1	F	R	65	2	3.0	29	0.0	1392	0.764	1.261	NaN	3969	3406

Row: individual data point

Column: particular dimension or feature

Matrices

A matrix is an $n \times d$ array of real numbers:

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nd} \end{pmatrix}$$

Notation: $X \in \mathbb{R}^{n \times d}$

Matrices

A matrix is an $n \times d$ array of real numbers:

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nd} \end{pmatrix}$$

Notation: $X \in \mathbb{R}^{n \times d}$

A **data matrix** is n data points, each with d features

Row and Column Vectors

Given an $n \times d$ matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ a_{21} & a_{22} & \cdots & a_{2d} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nd} \end{pmatrix}$$

Row vectors: $a_{i\bullet} = \begin{pmatrix} a_{i1} & a_{i2} & \cdots & a_{id} \end{pmatrix}$

Column vectors: $a_{\bullet j} = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{pmatrix}$

Matrix Addition

We can add two matrices of the same size: $A, B \in \mathbb{R}^{n \times d}$

$$A + B = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ a_{21} & a_{22} & \cdots & a_{2d} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nd} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1d} \\ b_{21} & b_{22} & \cdots & b_{2d} \\ \vdots & & & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nd} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1d} + b_{1d} \\ a_{21} + b_{12} & a_{22} + b_{22} & \cdots & a_{2d} + b_{2d} \\ \vdots & & & \vdots \\ a_{n1} + b_{n1} & a_{n2} + b_{n2} & \cdots & a_{nd} + b_{nd} \end{pmatrix}$$

Matrix-Scalar Multiplication

We can multiply a matrix $X \in \mathbb{R}^{n \times d}$ by a scalar $s \in \mathbb{R}$:

$$sA = s \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ a_{21} & a_{22} & \cdots & a_{2d} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nd} \end{pmatrix} = \begin{pmatrix} sa_{11} & sa_{12} & \cdots & sa_{1d} \\ sa_{21} & sa_{22} & \cdots & sa_{2d} \\ \vdots & & & \vdots \\ sa_{n1} & sa_{n2} & \cdots & sa_{nd} \end{pmatrix}$$

Matrix-Vector Multiplication

We can multiply an $n \times d$ matrix A with a d vector v:

$$Av = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ a_{21} & a_{22} & \cdots & a_{2d} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nd} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_d \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{n} a_{1j}v_j \\ \sum_{j=1}^{n} a_{2j}v_j \\ \vdots \\ \sum_{j=1}^{n} a_{nj}v_j \end{pmatrix}$$

The result is a d vector.

Matrix-Vector Multiplication

We can multiply an $n \times d$ matrix A with a d vector v:

$$Av = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ a_{21} & a_{22} & \cdots & a_{2d} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nd} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_d \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{n} a_{1j} v_j \\ \sum_{j=1}^{n} a_{2j} v_j \\ \vdots \\ \sum_{j=1}^{n} a_{nj} v_j \end{pmatrix}$$

The result is a d vector.

Each entry is a dot product between a row of A and v:

$$Av = \begin{pmatrix} \langle a_{1\bullet}, v \rangle \\ \langle a_{2\bullet}, v \rangle \\ \vdots \\ \langle a_{n\bullet}, v \rangle \end{pmatrix}$$

Matrices as Transformations

Consider a 2D matrix and coordinate vectors in $\mathbb{R}^2$:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \qquad v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Matrices as Transformations

Consider a 2D matrix and coordinate vectors in $\mathbb{R}^2$:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \qquad v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Then Av_1 and Av_2 result in the columns of A:

$$Av_1 = \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix}, \qquad Av_2 = \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix}$$

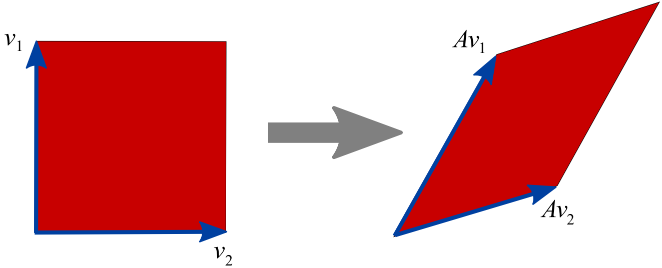
Matrices as Transformations

Consider a 2D matrix and coordinate vectors in $\mathbb{R}^2$:

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \qquad v_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Then Av_1 and Av_2 result in the columns of A:

$$Av_1 = \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix}, \qquad Av_2 = \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix}$$



Linearity of Matrix Multiplication

Given a matrix $A \in \mathbb{R}^{n \times d}$, vectors $v, w \in \mathbb{R}^d$, and scalars $s, t \in \mathbb{R}$:

$$A(sv + tw) = sAv + tAw$$

Matrix-Matrix Multiplication

Given matrices $A \in \mathbb{R}^{m \times d}$ and $B \in \mathbb{R}^{d \times n}$,
their **product** is:

$$(AB)_{ij} = \sum_{k=1}^{d} a_{ik} b_{kj}$$

Matrix-Matrix Multiplication

Given matrices $A \in \mathbb{R}^{m \times d}$ and $B \in \mathbb{R}^{d \times n}$, their **product** is:

$$(AB)_{ij} = \sum_{k=1}^{d} a_{ik} b_{kj}$$

Or, using dot products:

$$(AB)_{ij} = \langle a_{i\bullet}, b_{\bullet j} \rangle$$

Matrix-Matrix Multiplication

Given matrices $A \in \mathbb{R}^{m \times d}$ and $B \in \mathbb{R}^{d \times n}$, their **product** is:

$$(AB)_{ij} = \sum_{k=1}^{d} a_{ik} b_{kj}$$

Or, using dot products:

$$(AB)_{ij} = \langle a_{i\bullet}, b_{\bullet j} \rangle$$

Note: A column dimension = B row dimension

Matrix-Matrix Multiplication

Given matrices $A \in \mathbb{R}^{m \times d}$ and $B \in \mathbb{R}^{d \times n}$, their **product** is:

$$(AB)_{ij} = \sum_{k=1}^{d} a_{ik} b_{kj}$$

Or, using dot products:

$$(AB)_{ij} = \langle a_{i\bullet}, b_{\bullet j} \rangle$$

Note: A column dimension = B row dimension
Note: resulting product AB is $m \times n$

Identity Matrix

The identity matrix, denoted $I \in \mathbb{R}^{d \times d}$, is 1 on the diagonal and 0 off the diagonal:

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

Identity Matrix

The identity matrix, denoted $I \in \mathbb{R}^{d \times d}$, is 1 on the diagonal and 0 off the diagonal:

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

Multiplication by identity doesn't change the other matrix:

$AI = A,$ or $IB = B,$ for any $A \in \mathbb{R}^{n \times d}, B \in \mathbb{R}^{d \times n}$

Identity Matrix

The identity matrix, denoted $I \in \mathbb{R}^{d \times d}$, is 1 on the diagonal and 0 off the diagonal:

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

Multiplication by identity doesn't change the other matrix:

$$AI = A, \quad \text{or} \quad IB = B, \qquad \text{for any } A \in \mathbb{R}^{n \times d}, B \in \mathbb{R}^{d \times n}$$

Same thing for vectors:

$$Iv = v, \qquad \text{for any } v \in \mathbb{R}^d$$

Multilinear Regression

What if x_i has two features (x_{i1}, x_{i2})?

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \epsilon_i$$

Multilinear Regression

What if x_i has two features (x_{i1}, x_{i2})?

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \epsilon_i$$

Written as matrix-vector multiplication:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{n1} & x_{n2} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Multilinear Regression

What if x_i has two features (x_{i1}, x_{i2})?

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \epsilon_i$$

Written as matrix-vector multiplication:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{n1} & x_{n2} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Or, in matrix notation:

$$y = X\beta + \epsilon$$

Multilinear Regression: d features

What if x_i has d features $(x_{i1}, x_{i2}, \ldots, x_{id})$?

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{id}\beta_d + \epsilon_i$$

Multilinear Regression: d features

What if x_i has d features $(x_{i1}, x_{i2}, \ldots, x_{id})$?

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{id}\beta_d + \epsilon_i$$

Written as matrix-vector multiplication:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nd} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_d \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Multilinear Regression: d features

What if x_i has d features $(x_{i1}, x_{i2}, \ldots, x_{id})$?

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{id}\beta_d + \epsilon_i$$

Written as matrix-vector multiplication:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nd} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_d \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Or, in matrix notation:

$$y = X\beta + \epsilon$$

Multilinear Regression: Adding an Intercept

Use β_1 for the intercept:

$$y_i = \beta_1 + x_{i2}\beta_2 + \cdots + x_{id}\beta_d + \epsilon_i$$

Multilinear Regression: Adding an Intercept

Use β_1 for the intercept:

$$y_i = \beta_1 + x_{i2}\beta_2 + \cdots + x_{id}\beta_d + \epsilon_i$$

First column of X is ones:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{12} & \cdots & x_{1d} \\ 1 & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n2} & \cdots & x_{nd} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_d \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Multilinear Regression: Adding an Intercept

Use β_1 for the intercept:

$$y_i = \beta_1 + x_{i2}\beta_2 + \cdots + x_{id}\beta_d + \epsilon_i$$

First column of X is ones:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{12} & \cdots & x_{1d} \\ 1 & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n2} & \cdots & x_{nd} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_d \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}$$

Or, in matrix notation:

$$y = X\beta + \epsilon$$

Matrix Inverse

Given a square matrix $A \in \mathbb{R}^{d \times d}$, it's **inverse** is a matrix A^{-1} such that:

$$A^{-1}A = AA^{-1} = I$$

The inverse exists if and only if A has linearly independent columns.

Transpose

The transpose of a matrix $A \in \mathbb{R}^{n \times d}$ is a matrix $A^T \in \mathbb{R}^{d \times n}$ that “flips” row and column indices:

$$(A^T)_{ij} = A_{ji}$$

Transpose

The transpose of a matrix $A \in \mathbb{R}^{n \times d}$ is a matrix $A^T \in \mathbb{R}^{d \times n}$ that “flips” row and column indices:

$$(A^T)_{ij} = A_{ji}$$

Note: we can do inner products with transpose:

$$\langle v, w \rangle = v^T w$$

Least Squares Problem

Regression equation:

$$y = X\beta + \epsilon$$

Minimize the sum-of-squared error:

$$\begin{aligned}
\mathrm{SSE}(\beta) &= \|\epsilon\|^2 \\
&= \|y - X\beta\|^2 \\
&= (y - X\beta)^T (y - X\beta)
\end{aligned}$$

LS Solution

Derivative:

$$\frac{\partial}{\partial \beta}\mathrm{SSE}(\beta) = -X^T(y - X\beta) = -X^T y + X^T X\beta$$

LS Solution

Derivative:

$$\frac{\partial}{\partial \beta}\mathrm{SSE}(\beta) = -X^T(y - X\beta) = -X^T y + X^T X\beta$$

$$\begin{aligned}
X^T X\beta &= X^T y && \text{set derivative to zero} \\
(X^T X)^{-1}(X^T X)\beta &= (X^T X)^{-1} X^T y && \text{multiply by } (X^T X)^{-1} \\
\hat{\beta} &= (X^T X)^{-1} X^T y && \text{solve for } \beta
\end{aligned}$$