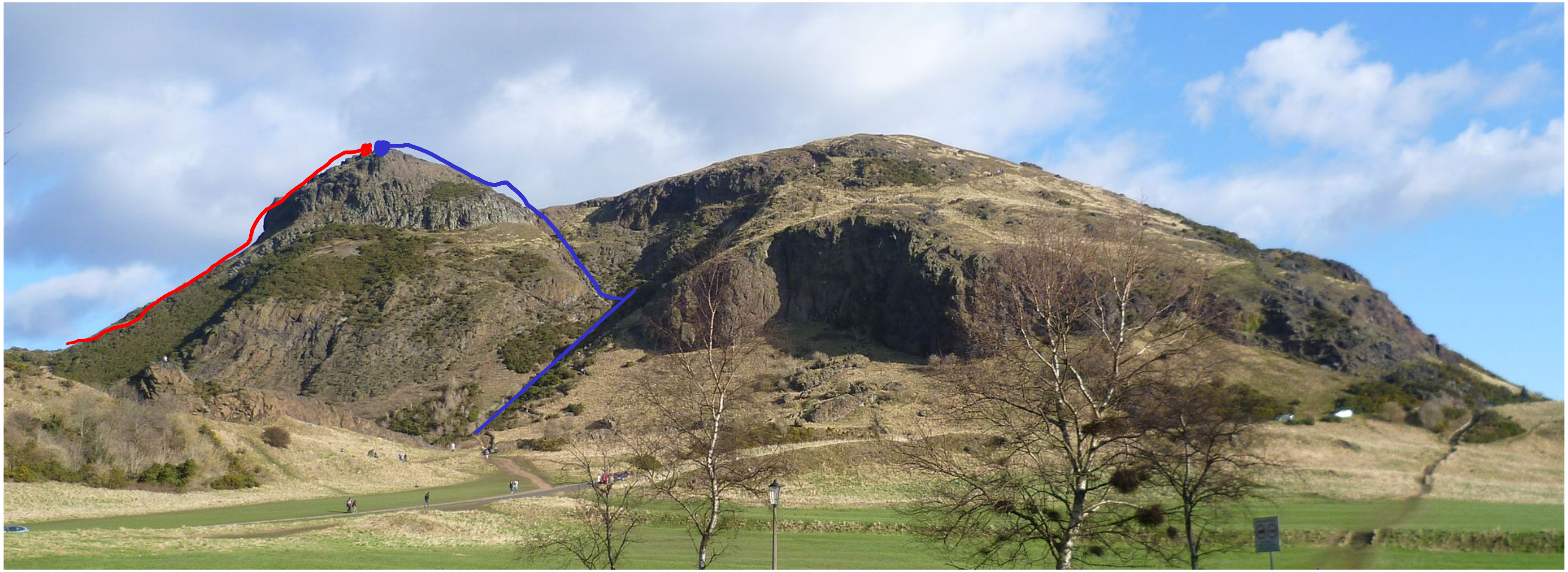
Foundations of Data Science:
Multiple regression -
Interpreting the coefficient equation



Kim T r ayno

Part 1 = simplifying

Part 2 = Coefficients themselves

Next video = properties of equation for coeffs.

- Linear algebra
- Calculus
- Definitions

$$\underline{\hat{\beta}} = \overbrace{(X^T X)}^{\text{normal mat.}}{}^{-1} \overbrace{X^T \underline{y}}^{\text{moment matrix}}$$

$$\underline{\hat{\beta}} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} \qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}^{(1)} - \hat{\beta}_2 \bar{x}^{(2)}$$

variance of $x^{(1)}$

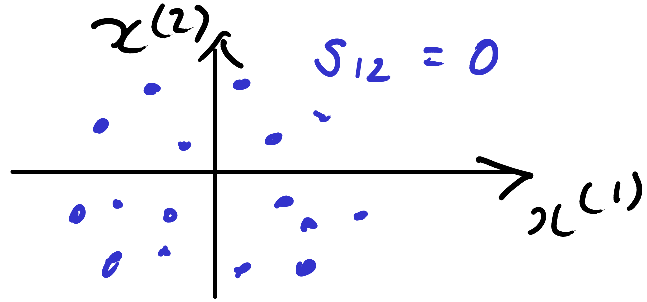
s_{12} = covariance of $x^{(1)}$ & $x^{(2)}$

2. $$S = \begin{pmatrix} s_1^2 & s_{12} \\ s_{12} & s_2^2 \end{pmatrix} = \frac{1}{n-1} X^T X$$

1. $$S = \begin{pmatrix} s_1^2 & 0 \\ 0 & s_2^2 \end{pmatrix}$$

Case 1.

$$S = \begin{pmatrix} s_1^2 & 0 \\ 0 & s_2^2 \end{pmatrix}$$



$$(X^T X)^{-1} = \frac{1}{n-1} S^{-1} = \frac{1}{n-1} \begin{pmatrix} 1/s_1^2 & 0 \\ 0 & 1/s_2^2 \end{pmatrix}$$

$$X^T \underline{y} = \begin{pmatrix} \sum_i x_{i1} y_i \\ \sum_i x_{i2} y_i \end{pmatrix} = (n-1) \begin{pmatrix} s_{1y} \\ s_{2y} \end{pmatrix}$$

covariance of $x^{(2)}$ and y

$s_{1y} = r_{1y} s_1 s_y$

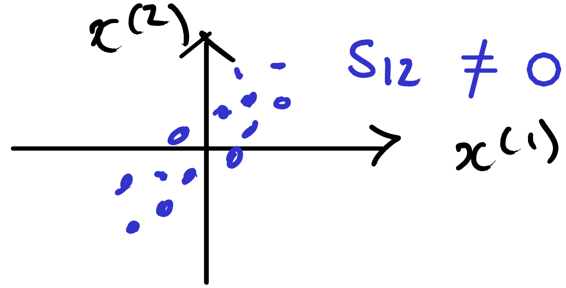
$$\Rightarrow \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} s_{1y}/s_1^2 \\ s_{2y}/s_2^2 \end{pmatrix} = \begin{pmatrix} r_{1y}\, s_y/s_1 \\ r_{2y}\, s_y/s_2 \end{pmatrix}$$

correlation of $x^{(1)}$ and y

c.f simple Linear Regression

Case 2

$$S = \begin{pmatrix} S_1^2 & S_{12} \\ S_{12} & S_2^2 \end{pmatrix}$$



$$= \begin{pmatrix} s_1^2 & r_{12} s_1 s_2 \\ r_{12} s_1 s_2 & s_2^2 \end{pmatrix}$$

$$\Rightarrow (X^T X)^{-1} = \frac{1}{n-1} S^{-1} = \frac{1}{(n-1)} \frac{1}{(1-r_{12}^2)} \begin{pmatrix} 1/s_1^2 & -r_{12}/(s_1 s_2) \\ -r_{12}/(s_1 s_2) & 1/s_2^2 \end{pmatrix}$$

$$X^T \underline{y} = (n-1) \begin{pmatrix} r_{1y} s_1 s_y \\ r_{2y} s_2 s_y \end{pmatrix}$$

$$\Rightarrow (X^T X)^{-1} X^T y = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \frac{1}{1-r_{12}^2} \begin{pmatrix} r_{1y} s_y / s_1 - r_{12} r_{2y} s_y / s_1 \\ r_{2y} s_y / s_2 - r_{12} r_{1y} s_y / s_2 \end{pmatrix}$$

$$= \frac{1}{1-r_{12}^2} \begin{pmatrix} (r_{1y} - r_{12} r_{2y}) s_y / s_1 \\ (r_{2y} - r_{12} r_{1y}) s_y / s_2 \end{pmatrix}$$

Interpretation & Example

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \frac{1}{1 - r_{12}^2} \begin{pmatrix} (r_{1y} - r_{12} r_{2y}) s_y / s_1 \\ (r_{2y} - r_{12} r_{1y}) s_y / s_1 \end{pmatrix}$$

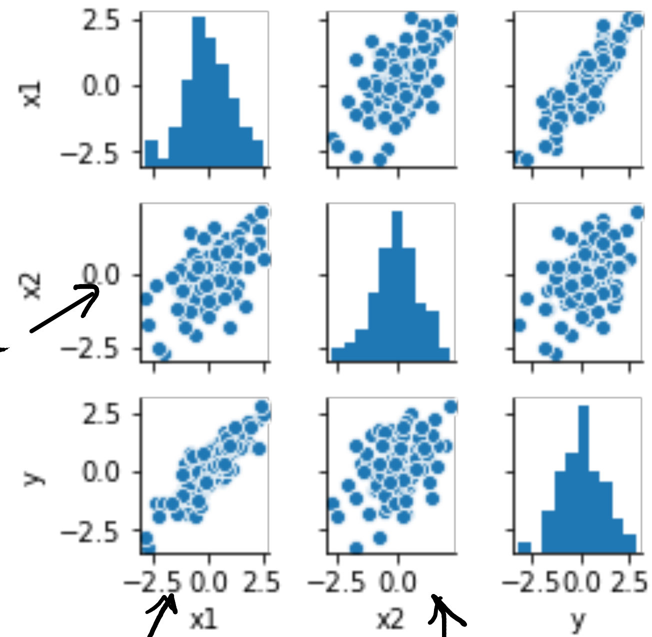
1. $r_{12} = 0 \Rightarrow$ Reduces to previous case

2. Eg.

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \frac{1}{1 - 0.5^2} \begin{pmatrix} 0.8 - 0.5 \times 0.4 \\ 0.4 - 0.5 \times 0.8 \end{pmatrix}$$

$$= \begin{pmatrix} 0.8 \\ 0 \end{pmatrix}$$

$r_{12} = 0.5$



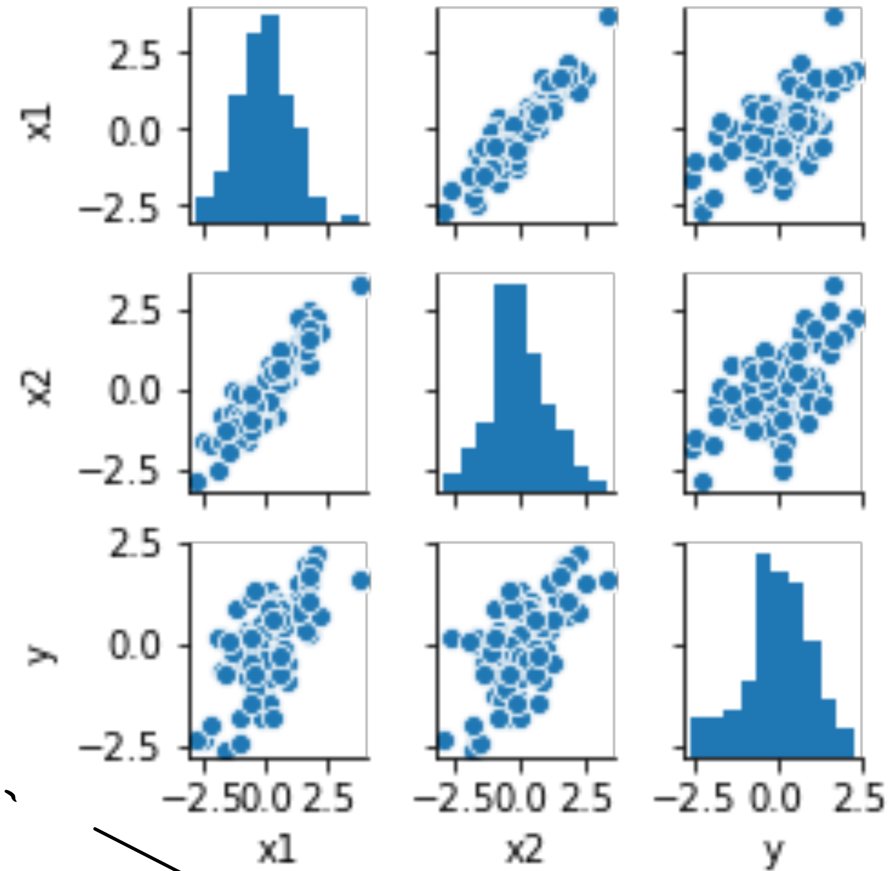
$r_{1y} = 0.8$ $\quad$ $r_{2y} = 0.4$

	coef	
Intercept	0.0130	
x1	0.9294	$\hat{\beta}_1$
x2	-0.0579	$\hat{\beta}_2$

Large correlations

	x1	x2	y
x1	1.000000	0.884385	0.565936
x2	0.884385	1.000000	0.576941
y	0.565936	0.576941	1.000000

$$\frac{0.01}{0.56} \simeq 2\%$$



$$\frac{0.11}{0.27} \sim 40\%$$

	coef	
Intercept	0.0030	
x1	0.2709	$\hat{\beta}_1$
x2	0.3803	$\hat{\beta}_2$

Colinearity

$r_{12} = 1$ $\qquad x^{(1)} = c\, x^{(2)}$ $\qquad c$ constant

$\frac{1}{1 - r_{12}^2} \to 0$

$\det(X^T X) = 0 \Rightarrow X^T X$ is singular

$\Rightarrow$ Error in fitting routine.