STAT 88: Lecture 36

Contents

Warm up: (Exercise 11.6.3)

Sometimes data scientists want to fit a linear model that has no intercept term. For example, this might be the case when the data are from a scientific experiement in which the attribute X can have values near 0 and there is a physical reason why the response Y must be 0 when $X = 0$.

So let (X, Y) be a random pair and suppose you want to predict Y by an estimator of the form aX for some a. Find the least squares predictor $\widehat{Y}$ among all predictors of this form.

Last time

Least squares regression

Let (X, Y) be a random pair. We write

- $E(X) = \mu_X$, $\mathrm{SD}(X) = \sigma_X$.
- $E(Y) = \mu_Y$, $\mathrm{SD}(Y) = \sigma_Y$.
- Correlation

unitless Covariance

$$r = \frac{E((X - \mu_X)(Y - \mu_Y))}{\sigma_X \sigma_Y}.$$

We wish to find the best fitting line $\widehat{Y} = \widehat{a}X + \widehat{b}$, through the scatter plot at all (X, Y) pairs. We showed that

$$\widehat{a} = r\frac{\sigma_Y}{\sigma_X} \text{ and } \widehat{b} = \mu_Y - \widehat{a} \cdot \mu_X.$$

$\hat{a}, \hat{b}$ minimize $MSE(a,b) = E((Y - (aX+b))^2)$

11.4. Bounds on Correlation

For a random pair (X, Y), the correlation is defined as

$$r = r(X,Y) = \frac{E((X-\mu_X)(Y-\mu_Y))}{\sigma_X \sigma_Y} = E\left(\frac{X-\mu_X}{\sigma_X} \cdot \frac{Y-\mu_Y}{\sigma_Y}\right) = E(X^\star Y^\star),$$

where $X^\star$ and $Y^\star$ are standardizations of X and Y respectively.

Our goal is to show that

$$-1 \le r \le 1.$$

As a preliminary, find $E(X^\star)$, $\mathrm{Var}(X^\star)$, and $E(X^{\star 2})$.

Lower Bound We will show that $r = E(X^\star Y^\star) \ge -1$.

Upper Bound Similarly,

Other Properties We can show

(a) $r(X,Y) = r(Y,X)$.

(b) $r(aX+b, cY+d) = \begin{cases} r(X,Y) & \text{if } ac > 0 \\ -r(X,Y) & \text{if } ac < 0 \end{cases}$

Example: (Exercise 11.6.7) Let (X, Y) be a random pair and let $r = r(X, Y)$. Let $X^\star$ be X in standard units and let $Y^\star$ be Y in standard units.

(a) Find $r(X^\star, Y^\star)$.

(b) Write the equation for $\widehat{Y}^\star$, the least squares linear predictor of $Y^\star$, based on $X^\star$.

11.5. The Error in Regression

The error in the regression estimate is called the residual and is defined as

$$D = Y - \widehat{Y}.$$

It is useful to write this in terms of the deviations $D_X = X - \mu_X$ and $D_Y = Y - \mu_Y$.

$$\widehat{Y} = \widehat{a}X + \widehat{b} = \widehat{a}X + \mu_Y - \widehat{a}\mu_X = \widehat{a}(X - \mu_X) + \mu_Y.$$

So,

$$\begin{aligned} D &= Y - \widehat{Y} \\ &= Y - [\widehat{a}(X - \mu_X) + \mu_Y] \\ &= Y - \mu_Y - \widehat{a}(X - \mu_X) \\ &= D_Y - \widehat{a}D_X. \end{aligned}$$

What is $E(D)$?

Mean Squared Error of Regression

The mean squared error of regression is $E((Y - \widehat{Y})^2) = E(D^2)$. Since $E(D) = 0$, we have $\mathrm{Var}(D) = E(D^2)$. Recall $\widehat{a} = r\frac{\sigma_Y}{\sigma_X}$ and $E(D_X D_Y) = r\sigma_X\sigma_Y$.

Let's find $\mathrm{Var}(D)$:

$$\begin{aligned} \mathrm{Var}(D) &= E(D^2) \\ &= E(D_Y^2) - 2\widehat{a}E(D_X D_Y) + \widehat{a}^2 E(D_X^2) \\ &= \sigma_Y^2 - 2r\frac{\sigma_Y}{\sigma_X} r\sigma_X\sigma_Y + r^2\frac{\sigma_Y^2}{\sigma_X^2}\sigma_X^2 \\ &= \sigma_Y^2 - 2r^2\sigma_Y^2 + r^2\sigma_Y^2 \\ &= \sigma_Y^2 - r^2\sigma_Y^2 \\ &= (1 - r^2)\sigma_Y^2. \end{aligned}$$

So

$$\mathrm{SD}(D) = \sqrt{1 - r^2}\,\sigma_Y.$$

r As a Measure of Linear Association Note that

$$E(D) = 0 \text{ and } \mathrm{SD}(D) = \sqrt{1 - r^2}\sigma_Y.$$

So if r is close to ± 1, $\mathrm{SD}(D)$ is close to 0, which implies that Y is close to $\widehat{Y}$. In other words, Y is close to being a linear function of X.

In the extreme case $r = \pm 1$, $\mathrm{SD}(D) = 0$ and Y is a perfectly linear function of X.

The Residual is Uncorrelated with X We will show that the correlation between X and residual D is zero. Note that

$$r(D, X) = \frac{E((D - \mu_D)(X - \mu_X))}{\sigma_D \sigma_X} = \frac{1}{\sigma_D \sigma_X} E(D D_X),$$

because $\mu_D = 0$. We thus show $E(DD_X) = 0$:

$$\begin{aligned} E(DD_X) &= E((D_Y - \widehat{a} D_X) D_X) \\ &= E(D_X D_Y) - \widehat{a} E(D_X^2) \\ &= r \sigma_X \sigma_Y - r \frac{\sigma_Y}{\sigma_X} \sigma_X^2 \\ &= 0. \end{aligned}$$