Partial regression plot

October 29, 2018

Theorem

Let $\hat{\beta}$ be the vector of coefficients from the regression of Y on X, with residuals $\hat{e}$, and $X^{(j)}$ the j-th column of X. Let $b^{(j)}$ be the slope from the simple linear regression of $Y^{(j)} := Res(Y, X(j))$ on $X^{(j)} := Res(X(j), X^{-j})$, with residuals $e^{(j)}$. Then $b^{(j)} = \hat{\beta}_j$.

Theorem

Let $\hat{\beta}$ be the vector of coefficients from the regression of Y on X, with residuals $\hat{e}$, and $X^{(j)}$ the j-th column of X. Let $b^{(j)}$ be the slope from the simple linear regression of $Y^{(j)} := Res(Y, X(j))$ on $X^{(j)} := Res(X(j), X^{-j})$, with residuals $e^{(j)}$. Then $b^{(j)} = \hat{\beta}_j$.

-
- We prove the case $j = p$; the generalization is obvious.

Theorem

Let $\hat{\beta}$ be the vector of coefficients from the regression of Y on X, with residuals $\hat{e}$, and $X^{(j)}$ the j-th column of X. Let $b^{(j)}$ be the slope from the simple linear regression of $Y^{(j)} := Res(Y, X(j))$ on $X^{(j)} := Res(X(j), X^{-j})$, with residuals $e^{(j)}$. Then $b^{(j)} = \hat{\beta}_j$.

-
- We prove the case $j = p$; the generalization is obvious.
- First, let's consider the simple linear regression of $Y^{(p)}$ on $X^{(p)}$.

Theorem

Let $\hat{\beta}$ be the vector of coefficients from the regression of Y on X, with residuals $\hat{e}$, and $X^{(j)}$ the j-th column of X. Let $b^{(j)}$ be the slope from the simple linear regression of $Y^{(j)} := Res(Y, X(j))$ on $X^{(j)} := Res(X(j), X^{-j})$, with residuals $e^{(j)}$. Then $b^{(j)} = \hat{\beta}_j$.

-
- We prove the case $j = p$; the generalization is obvious.
- First, let's consider the simple linear regression of $Y^{(p)}$ on $X^{(p)}$.
- Let $H(-p)$ be the projection matrix on all columns in X except $X(p)$, the pth column.

Theorem

Let $\hat{\beta}$ be the vector of coefficients from the regression of Y on X, with residuals $\hat{e}$, and $X^{(j)}$ the j-th column of X. Let $b^{(j)}$ be the slope from the simple linear regression of $Y^{(j)} := Res(Y, X(j))$ on $X^{(j)} := Res(X(j), X^{-j})$, with residuals $e^{(j)}$. Then $b^{(j)} = \hat{\beta}_j$.

-
- We prove the case $j = p$; the generalization is obvious.
- First, let's consider the simple linear regression of $Y^{(p)}$ on $X^{(p)}$.
- Let $H(-p)$ be the projection matrix on all columns in X except $X(p)$, the pth column.
- Then $Y^{(p)} = (I - H(-p))Y$ and $X^{(p)} = (I - H(-p))X(p)$.

Theorem

Let $\hat{\beta}$ be the vector of coefficients from the regression of Y on X, with residuals $\hat{e}$, and $X^{(j)}$ the j-th column of X. Let $b^{(j)}$ be the slope from the simple linear regression of $Y^{(j)} := Res(Y, X(j))$ on $X^{(j)} := Res(X(j), X^{-j})$, with residuals $e^{(j)}$. Then $b^{(j)} = \hat{\beta}_j$.

-
- We prove the case $j = p$; the generalization is obvious.
- First, let's consider the simple linear regression of $Y^{(p)}$ on $X^{(p)}$.
- Let $H(-p)$ be the projection matrix on all columns in X except $X(p)$, the pth column.
- Then $Y^{(p)} = (I - H(-p))Y$ and $X^{(p)} = (I - H(-p))X(p)$.
- Since both $Y^{(p)}$ and $X^{(p)}$ are residuals from a fitted linear model with an intercept, their entries have sample mean 0; therefore this is simple linear regression through the origin.

- Hence:

$$b^{(p)} = \frac{(X^{(p)})^T Y^{(p)}}{\|X^{(p)}\|_2^2}.$$

- Hence:

$$b^{(p)} = \frac{(X^{(p)})^T Y^{(p)}}{\|X^{(p)}\|_2^2}.$$

- And so

$$b^{(p)} = \frac{Y^T(I - H(-p))X(p)}{X(p)^T(I - H(-p))X(p)}.$$

- Hence:

$$b^{(p)} = \frac{(X^{(p)})^T Y^{(p)}}{\|X^{(p)}\|_2^2}.$$

- And so

$$b^{(p)} = \frac{Y^T(I - H(-p))X(p)}{X(p)^T(I - H(-p))X(p)}.$$

- We need to prove that $\hat{\beta}_p$ is also equal to this quantity.

- Hence:

$$b^{(p)} = \frac{(X^{(p)})^T Y^{(p)}}{\|X^{(p)}\|_2^2}.$$

- And so

$$b^{(p)} = \frac{Y^T(I - H(-p))X(p)}{X(p)^T(I - H(-p))X(p)}.$$

- We need to prove that $\hat{\beta}_p$ is also equal to this quantity.
- To do so, consider fitting a linear regression under the constraint $\beta_p = b$ for some fixed value b; the corresponding minimization problem would be the following:

- Hence:

$$b^{(p)} = \frac{(X^{(p)})^T Y^{(p)}}{\|X^{(p)}\|_2^2}.$$

- And so

$$b^{(p)} = \frac{Y^T(I - H(-p))X(p)}{X(p)^T(I - H(-p))X(p)}.$$

- We need to prove that $\hat{\beta}_p$ is also equal to this quantity.
- To do so, consider fitting a linear regression under the constraint $\beta_p = b$ for some fixed value b; the corresponding minimization problem would be the following:
-
$$\min_{\alpha \in \mathbb{R}^p} \|Y - V\alpha - X(p)b\|_2^2$$

  where V contains the first p columns of X. Let $\alpha(b)$ denote the minimizer of the above problem.

- Note that $\alpha(0)$ is the least squares regression of Y on V, while $\alpha(\hat{\beta}_p)$ represents the first p coefficients of the full linear regression of Y on X.

- Note that $\alpha(0)$ is the least squares regression of Y on V, while $\alpha(\hat{\beta}_p)$ represents the first p coefficients of the full linear regression of Y on X.
- We want a formula for $\alpha(b)$ for all b; therefore we take the gradient in to obtain the stationary condition.

- Note that $\alpha(0)$ is the least squares regression of Y on V, while $\alpha(\hat{\beta}_p)$ represents the first p coefficients of the full linear regression of Y on X.
- We want a formula for $\alpha(b)$ for all b; therefore we take the gradient in to obtain the stationary condition. Which, upon solving for, yields:

$$\alpha(b) \;=\; (V^T V)^{-1} V^T (Y - X(p)b).$$

- Note that $\alpha(0)$ is the least squares regression of Y on V, while $\alpha(\hat{\beta}_p)$ represents the first p coefficients of the full linear regression of Y on X.
- We want a formula for $\alpha(b)$ for all b; therefore we take the gradient in to obtain the stationary condition. Which, upon solving for, yields:

$$\alpha(b) \;=\; (V^T V)^{-1} V^T (Y - X(p)b).$$

- Note that $X\hat{\beta} = V\alpha(\hat{\beta}_p) \;+\; X(p)\,\hat{\beta}_p$; combined with the formula for $\alpha(b)$ this yields:

$$\begin{aligned} X\hat{\beta} &= V(V^T V)^{-1} V^T (Y - X(p)\hat{\beta}_p) \;+\; X(p)\hat{\beta}_p \\ &= H(-p)Y \;+\; (I - H(-p))X(p)\hat{\beta}_p \\ &= HY, \end{aligned}$$

where the last equality is the fundamental equality of least squares (here H is the projection matrix on all columns of X).

- Multiplying the above equalities through by $X(p)T$ yields:

$$\begin{aligned} X(p)^T\,Y &= X(p)^T\,H\,Y \\ &= X(p)^T H(-p)Y \;+\; X(p)^T(I - H(-p))X(p)\hat{\beta}_p, \end{aligned}$$

solving for $\hat{\beta}_p$ leads to the desired result.