Partial Least Squares Regression -- a simple two-dimensional example.
Instructor: Nam Sun Wang

Step 0. Generate X and Y Data. The first two independent variables $x^{<0>}$ and $x^{<1>}$ are somewhat dependent. Although $x^{<0>}$ is 2 times of $x^{<1>}$, this difference is absorbed during variance-scaling.

Number of points: $N := 50 \qquad i := 0..N$

Dimension: $m := 1 \qquad j := 0..m$

$$X^{<i>} := (\text{rnd}(1) - 0.5)\cdot\begin{pmatrix}2\\1\end{pmatrix} + (\text{rnd}(1) - 0.5)\cdot\begin{pmatrix}0\\1\end{pmatrix} + (\text{rnd}(1) - 0.5)\cdot\begin{pmatrix}2\\0\end{pmatrix} \qquad X := X^T$$

The dependent variable Y depends on $x^{<0>}$ and $x^{<1>}$ with a 2:1 ratio. The relative contribution to Y is proportional to the standard deviation x_{stdev} and the multiplicative coefficient a.

$$a := \begin{pmatrix}2\\1\end{pmatrix} \qquad y(x) := x\cdot a$$

$$Y_i := a_0\cdot X_{i,0} + a_1\cdot X_{i,1} + (\text{rnd}(0.1) - 0.05)$$

Step 1a. Mean-Centering:

$$x_{mean_j} := \text{mean}\left(X^{<j>}\right) \qquad x_{mean} := x_{mean}^T \qquad X^{<j>} := X^{<j>} - x_{mean_{0,j}}$$

$$x_{mean} = \begin{pmatrix}4.487\cdot 10^{-4} & -0.015\end{pmatrix}$$

$$y_{mean} := \text{mean}(Y) \qquad Y := Y - y_{mean} \qquad y_{mean} = -0.02$$

Step 1b. Variance-Scaling:

$$x_{stdev_j} := \text{stdev}\left(X^{<j>}\right) \qquad x_{stdev} := x_{stdev}^T \qquad X^{<j>} := \frac{X^{<j>}}{x_{stdev_{0,j}}}$$

$$x_{stdev} = \begin{pmatrix}0.913 & 0.358\end{pmatrix}$$

Step 2. Find the eigenvalues and eigenvectors of the mean-centered and variance-scaled covariance matrix $X^T\cdot Y\cdot Y^T\cdot X$.

Covariance matrix:

$$W := X^T\cdot Y\cdot Y^T\cdot X \qquad W = \begin{pmatrix}1.047\cdot 10^4 & 6.59\cdot 10^3\\ 6.59\cdot 10^3 & 4.149\cdot 10^3\end{pmatrix} \qquad |W| = 4.195\cdot 10^{-8}$$ ← Singular.

Eigenvalue/eigenvector

$$\lambda := \text{reverse}(\text{sort}(\text{eigenvals}(W))) \qquad \lambda^T = \begin{pmatrix}1.462\cdot 10^4 & 3.052\cdot 10^{-12}\end{pmatrix}$$

$$w^{<j>} := \text{eigenvec}\left(W, \lambda_j\right) \qquad w = \begin{pmatrix}0.846 & -0.533\\ 0.533 & 0.846\end{pmatrix}$$

Correct the weighting vector $w^{<0>}$.

$$w^{<0>} := w^{<0>}\cdot\frac{\left|X^T\cdot X\cdot w^{<0>}\right|}{\left|w^{<0>T}\cdot X^T\cdot X\cdot w^{<0>}\right|} \qquad w^{<0>} = \begin{pmatrix}0.855\\0.539\end{pmatrix} \qquad \left|w^{<0>}\right| = 1.011$$

Step 3. Score and loading vectors for X and Y.

score (column) vector for X: $t^{\langle 0\rangle} := X\cdot w^{\langle 0\rangle}$

loading (row) vector for Y: $q^{\langle 0\rangle} := Y^T\cdot t^{\langle 0\rangle}$ normalize: $q^{\langle 0\rangle} := \dfrac{q^{\langle 0\rangle}}{\left|q^{\langle 0\rangle}\right|}$ $q = 1$

score (column) vector for Y: $u^{\langle 0\rangle} := Y\cdot q^{\langle 0\rangle}$

loading (row) vector for X: $p^{\langle 0\rangle} := X^T\cdot t^{\langle 0\rangle}$ normalize: $p^{\langle 0\rangle} := \dfrac{p^{\langle 0\rangle}}{\left|p^{\langle 0\rangle}\right|}$ $p = \begin{pmatrix} 0.759 \\ 0.651 \end{pmatrix}$

A comparison of different vectors.

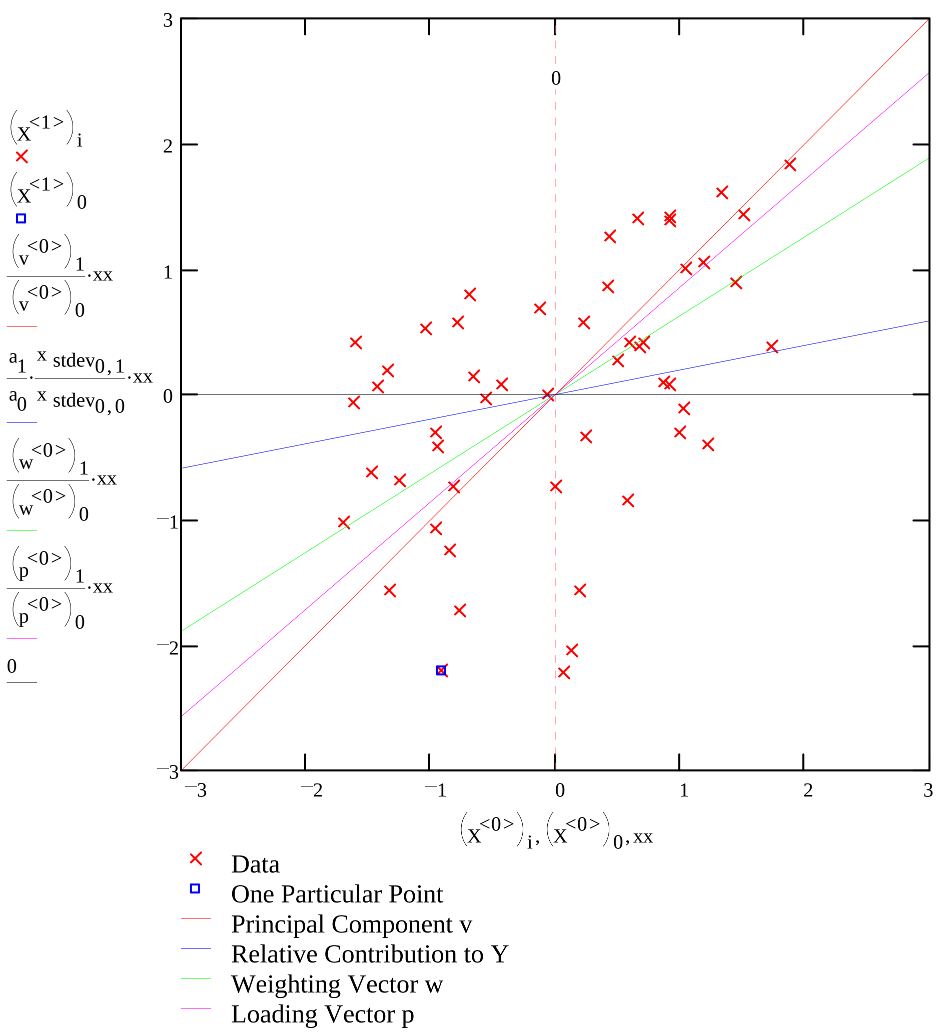
Eigenvalue/eigenvector of the covariance matrix $X^T\cdot X$ -- Principal Component Regression (PCR).

$V := X^T\cdot X$

$\lambda_v := \text{reverse}(\text{sort}(\text{eigenvals}(V)))$ $\lambda_v{}^T = (\ 76.142 \quad 25.858\ )$

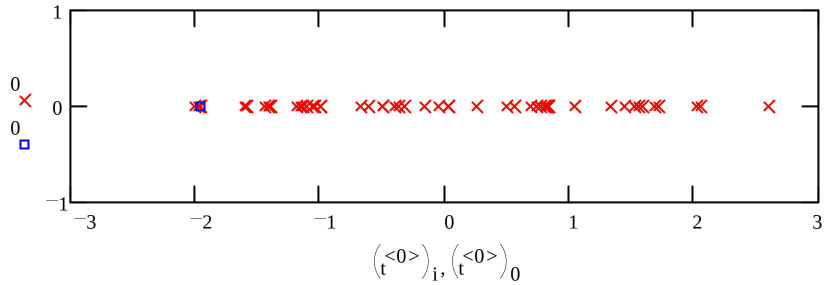
$v^{\langle j\rangle} := \text{eigenvec}\left(V, \lambda_{v_j}\right)$ $v = \begin{pmatrix} 0.707 & -0.707 \\ 0.707 & 0.707 \end{pmatrix}$

$xx := -3 .. 3$



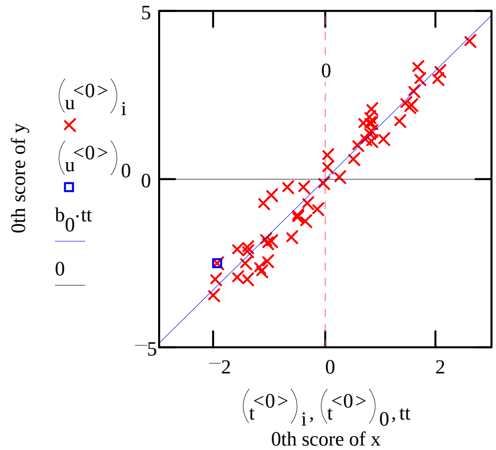
— axis

$t^{<0>}$, 1st score vector of X:



Step 4. Regression.

$b_0 := \mathrm{slope}\left(t^{\langle 0\rangle}, u^{\langle 0\rangle}\right)$ $\qquad b = 1.623$ $\qquad tt := -3, -2.9 .. 3$



Step 5. Compute the residual matrices

$$E := X - t^{\langle 0\rangle} \cdot p^{\langle 0\rangle\, T}$$

$$F := Y - t^{\langle 0\rangle} \cdot q^{\langle 0\rangle\, T} \cdot b_0$$

Step 6. Goodness of fit.

$sse_{old} := Y \cdot Y \qquad sse_{old} = 210.322$

$sse := F \cdot F \qquad sse = 11.942$

$r2 := \dfrac{sse_{old} - sse}{sse_{old}} \qquad r2 = 94.322 \cdot \%$

$r := \sqrt{r2} \qquad r = 97.119 \cdot \%$

2nd Iteration. We start with the residual independent variables E and F.

Covariance matrix: $W := E^T \cdot F \cdot F^T \cdot E \qquad W = \begin{pmatrix} 90.28 & -143.389 \\ -143.389 & 227.742 \end{pmatrix}$

Eigenvalue: $\lambda := \mathrm{reverse}(\mathrm{sort}(\mathrm{eigenvals}(W))) \qquad \lambda^T = (\ 318.022 \quad 0\ )$

Eigenvector: $w^{\langle 1\rangle} := \mathrm{eigenvec}\left(W, \lambda_0\right) \qquad w = \begin{pmatrix} 0.855 & -0.533 \\ 0.539 & 0.846 \end{pmatrix}$

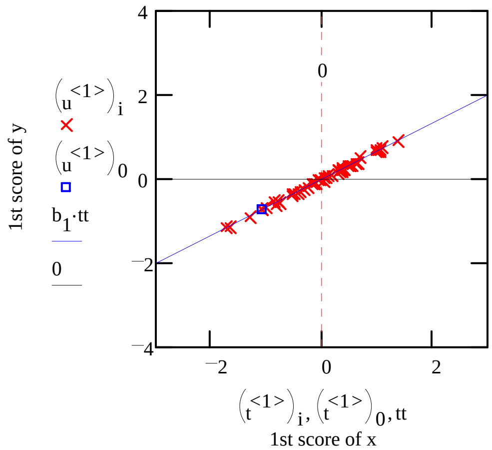
score (column) vector for X: $t^{\langle 1\rangle} := E \cdot w^{\langle 1\rangle}$

loading (row) vector for Y: $q^{\langle 1\rangle} := F^T \cdot t^{\langle 1\rangle}$ normalize: $q^{\langle 1\rangle} := \dfrac{q^{\langle 1\rangle}}{\left|q^{\langle 1\rangle}\right|} \qquad q^{\langle 1\rangle} = -1$

score (column) vector for Y: $u^{\langle 1\rangle} := F \cdot q^{\langle 1\rangle}$

loading (row) vector for X: $p^{\langle 1\rangle} := E^T \cdot t^{\langle 1\rangle}$ normalize: $p^{\langle 1\rangle} := \dfrac{p^{\langle 1\rangle}}{\left|p^{\langle 1\rangle}\right|} \qquad p^{\langle 1\rangle} = \begin{pmatrix} -0.533 \\ 0.846 \end{pmatrix}$

Regression: $b_1 := \text{slope}\left(t^{\langle 1\rangle}, u^{\langle 1\rangle}\right)$ $\quad b = \begin{pmatrix} 1.623 \\ 0.667 \end{pmatrix}$ $\quad tt := -3, -2.9 .. 3$



Residual matrices: $E := E - t^{\langle 1\rangle} \cdot p^{\langle 1\rangle T}$

$F := F - t^{\langle 1\rangle} \cdot q^{\langle 1\rangle T} \cdot b_1$

Goodness of fit: $sse := F \cdot F$ $\quad sse = 0.042$

$r2 := \dfrac{sse_{old} - sse}{sse_{old}}$ $\quad r2 = 99.98 \cdot \%$

$r := \sqrt{r2}$ $\quad r = 99.99 \cdot \%$

Step 6. Regression Model. (Be sure to take care of both mean-centering and variance-scaling)

Variance-scaling: $x_{stdev\_inv_{j,j}} := \dfrac{1}{x_{stdev_{0,j}}}$

$I_{j,j} := 1$ ... an identity matrix.

$$y_{regress}(x) := \left(x - x_{mean}\right) \cdot x_{stdev\_inv} \cdot \left[w^{\langle 0\rangle} \cdot q^{\langle 0\rangle T} \cdot b_0 + \left(I - w^{\langle 0\rangle} \cdot p^{\langle 0\rangle T}\right) \cdot w^{\langle 1\rangle} \cdot q^{\langle 1\rangle T} \cdot b_1 \right] ... + y_{mean}$$

Let us examine the slope and intercept with 2 terms (j=0,1).

slope:

$$x_{stdev\_inv} \cdot \left[w^{\langle 0\rangle} \cdot q^{\langle 0\rangle T} \cdot b_0 + \left(I - w^{\langle 0\rangle} \cdot p^{\langle 0\rangle T}\right) \cdot w^{\langle 1\rangle} \cdot q^{\langle 1\rangle T} \cdot b_1 \right] = \begin{pmatrix} 2.002 \\ 1.011 \end{pmatrix} \longleftrightarrow \quad a = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$$

Examples:

$y_{regress}((5 \;\; 0.5)) = 10.508$ $\longleftrightarrow$ $\quad y((5 \;\; 0.5)) = 10.5$ $\leftarrow$ O.K.

$y_{regress}((5 \;\; -0.5)) = 9.497$ $\longleftrightarrow$ $\quad y((5 \;\; -0.5)) = 9.5$ $\leftarrow$ O.K.

Miscellaneous Checks.

Check: orthogonality of p. (The vectors in p are not mutually orthogonal.)

$$p^T \cdot p = \begin{pmatrix} 1 & 0.146 \\ 0.146 & 1 \end{pmatrix}$$

Check: orthogonality of w. (The vectors in w are mutually orthogonal, but the 0th vector is not normalized.)

$$w^T \cdot w = \begin{pmatrix} 1.022 & 0 \\ 0 & 1 \end{pmatrix}$$

Check: $X = t \cdot p^T$

X =

	0
0	-0.907
1	-1.042
2	-0.434
3	-1.608

$t \cdot p^T$ =

	0
0	-0.907
1	-1.042
2	-0.434
3	-1.608

Check: $Y = \Sigma(t \cdot q^T \cdot b) + F$

Y =

	0
0	-2.479
1	-1.736
2	-0.715
3	-2.762
4	-1.997

$\sum_{j=0}^{1} t^{\langle j \rangle} \cdot q^{\langle j \rangle T} \cdot b_j + F$ =

	0
0	-2.479
1	-1.736
2	-0.715
3	-2.762
4	-1.997

Check: E=0

E =

	0	1
0	0	0
1	0	0
2	0	0
3	0	0
4	0	0