Linear Methods for Regression:
Methods using derived input directions

MOTIVATION:

● When a **large number** of (correlated) **variables** Xj, j=1,…,p are available, they may be **linearly combined** in a **small number** of **components** (projections) Zm, m=1,…,M, with M<=p.

● These **components** can be used as inputs in **regression**.

● Different **methods** are available for **constructing linear combinations** of variables

● Partial least squares

PRINCIPAL COMPONENT REGRESSION:

Linear components Zm are defined by **Principal Component**
Analysis (PCA).

Principal components (Karhunen-Loeve) directions of **X** are
computed by **SVD** of **X** (**eigenvalue decomposition** of **X**T**X,** if **X** is
standardized).

● The **SVD** of the N x p matrix **X** can be written as:

X = UDVT

where:

● U (N x p) and V (p x p) are **orthogonal** matrices

● Columns of U span the **column space** of X

● Columns of V span the **row space** of X

● D is a p x p diagonal matrix with entries d1 >= d2 >= … >= dp >=0
singular values of X.

PARAMETER LEARNING:

Principal Component Regression forms the **derived input columns**
zm=**X***vm
and then regresses **y** on **z**1, **z**2,…,**z**M, for some **M<=p**

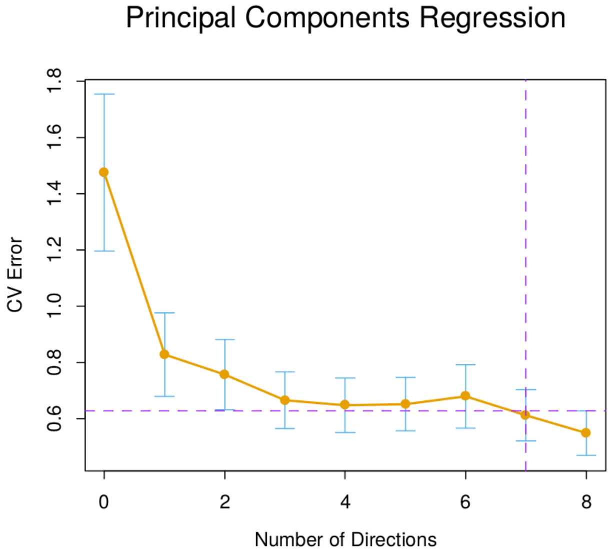
Since the zm are **orthogonal**, this regression is a sum of univariate
regressions:
where .
● Since the zm are linear combinations of the original xj, the
coefficients of variables **x**j can be written as

Data standardization is needed (as in ridge regression) since
principal components depend on variable scale.

- *If M=p then PCR corresponds to OLS since the columns of Z=UD*
span the column space of X.

Similarities between ridge regression and PCR:

- *Both operate on principal components of X*
- *Ridge shrinks more the components with small eigenvalues*
(directions with smaller variance)
- *PCR discards the p-M smallest eigenvalue components*



Similarities between ridge regression and PCR:

● **Both** operate on **principal components** of **X**

● **Ridge shrinks more** the components with **small eigenvalues**
(directions with **smaller variance**)

● **PCR discards** the p-M **smallest eigenvalue** components

Principal component index

Principal Component Regression forms the **derived input columns**
zm=**X***vm
and then regresses **y** on **z**1, **z**2,…,**z**M, for some **M<=p**

Since the zm are **orthogonal**, this regression is a sum of univariate regressions:

$$\hat{\mathbf{y}}^{\mathrm{pcr}}_{(M)} = \bar{y}\mathbf{1} + \sum_{m=1}^{M} \hat{\theta}_m \mathbf{z}_m,$$

Since the zm are linear combinations of the original xj, the
coefficients of variables **x**j can be written as

$$\hat{\beta}^{\text{pcr}}(M) = \sum_{m=1}^{M} \hat{\theta}_m v_m.$$

The **first principal components** direction v1 (**eigenvector** of **X**T**X**)
has the property that **z**1 = **X***v1 has the **largest sample variance**
amongst all normalized linear combinations of columns of **X**
Var(**z**1) = Var(**X***v1) = d1
2/N,
where d1 is the eigenvalue of **X**T**X**
with maximum absolute value and N
is the total number of observations.

● **Subsequent principal components**
zj have **maximum variance** and
are **orthogonal** to the earlier ones