Principal Component Regression (PCR): A Detailed Breakdown

Tanmay Chopra

November 21, 2025

1 What is PCR?

Principal Component Regression (PCR) is a regression technique that systematically combines two steps:

1. **PCA** (Principal Component Analysis) for dimensionality reduction and decorrelation of predictors.
2. **Linear Regression** performed on the derived principal components instead of the original features.

PCR is especially useful when:

- Features are highly correlated, leading to the **multicollinearity problem** in standard OLS.
- The number of features (p) is large compared to the number of samples (n).

PCR helps stabilize regression coefficients and often improves prediction performance by managing the variance-bias trade-off.

—

2 Intuition

The core idea is to replace the potentially unstable, correlated original features with a smaller, cleaner set of orthogonal components that still capture most of the variance in the data.

- **PCA step:** Transform the set of correlated predictors into a set of uncorrelated Principal Components (PCs).
- **Regression step:** Regress the target variable (y) on these PCs.
- **Benefit:** By using only the first few PCs (those with the most variance), we effectively reduce noise and avoid potential overfitting caused by unstable coefficients.

—

3 How PCR Works: Step by Step

Let X represent the predictor matrix and y represent the response vector.

3.1 Step 1: Standardize the Predictors

Since PCA is scale-sensitive, we standardize the predictor matrix X to Z:

$$Z = \frac{X - \mu}{\sigma}$$

3.2 Step 2: Perform PCA on X

Compute the covariance matrix of Z, find its eigenvectors and eigenvalues, and sort the Principal Components (PCs) by decreasing variance (eigenvalue magnitude).

3.3 Step 3: Select Top k Principal Components

Decide how many PCs, k, to retain.

- This is often done by retaining PCs that explain a high percentage of the cumulative variance (e.g., 90%–95%).
- Alternatively, k can be chosen via **cross-validation** to optimize prediction accuracy.

Let W_k be the matrix containing the top k eigenvectors (loadings).

3.4 Step 4: Transform the Data

Project the standardized predictor data Z onto the selected k PCs to obtain the component scores matrix T:

$$T = ZW_k$$

T is an $n \times k$ matrix of uncorrelated components.

3.5 Step 5: Regress y on T

Fit a linear regression model using the component scores T as predictors:

$$y = T\beta + \epsilon$$

The coefficient vector β is estimated via Ordinary Least Squares (OLS) as $\hat{\beta} = (T^T T)^{-1} T^T y$. This vector contains the regression coefficients in the PC space.

3.6 Step 6: Convert Back to Original Feature Space (Optional)

If interpretation in terms of the original features is required, the coefficients $\hat{\beta}_{\text{original}}$ can be calculated:

$$\hat{\beta}_{\text{original}} = W_k \hat{\beta}$$

—

4 Key Points About PCR

- **Addresses Multicollinearity:** By using PCs, which are orthogonal, the instability caused by highly correlated features is resolved.
- **Dimensionality Reduction:** Using $k < p$ PCs simplifies the model, which helps reduce overfitting and noise.
- **Variance-Focused:** PCR discards PCs with low variance. If a feature is important for predicting y but has low variance in X, PCR might inadvertently discard it.
- **Linear Method:** It assumes a linear relationship between the Principal Components and the response variable y.

—

5 PCR vs. PLS (Partial Least Squares)

Feature	PCR	PLS
Focus	Maximizes variance in X	Maximizes **covariance** between X
PCs used for regression	Unsupervised (ignores y)	Supervised (considers y)
Risk	May discard low-variance but predictive features	Less likely to discard predictive infor

PLS is generally preferred when the predictive power is the primary goal, as its components are constructed to optimize the relationship with the response y.

—

6 Quick Example

Suppose you have 10 highly correlated chemical features predicting a reaction yield.

1. Standardize the 10 features.
2. Perform PCA to get 10 PCs.
3. Keep top 3 PCs that explain 95% of variance.
4. Regress the reaction yield on these 3 PCs.
5. Use the resulting model for prediction.

This process successfully avoids instability due to multicollinearity, reduces noise, and yields a simpler, more robust model.

$$\text{PCR} = \text{PCA} + \text{Linear Regression}$$

It reduces the predictor space to orthogonal components and then builds a regression model, which is more stable in the presence of correlated predictors.