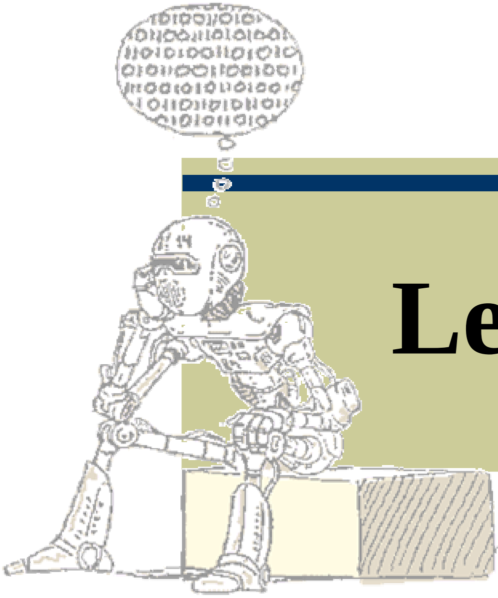


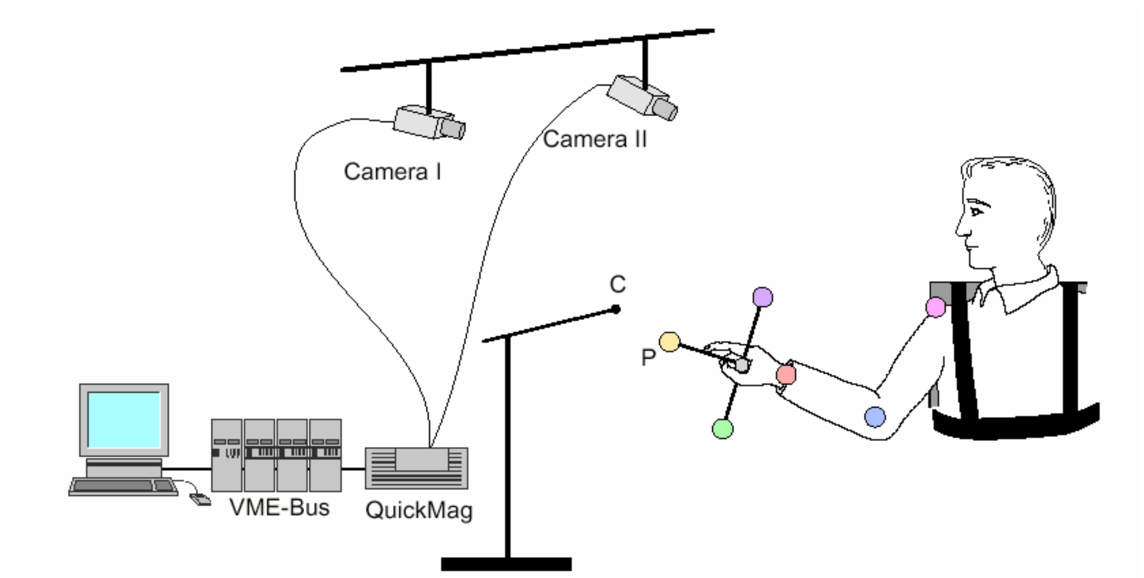
Lecture IX– Dim. Reduction (2)

Contents:

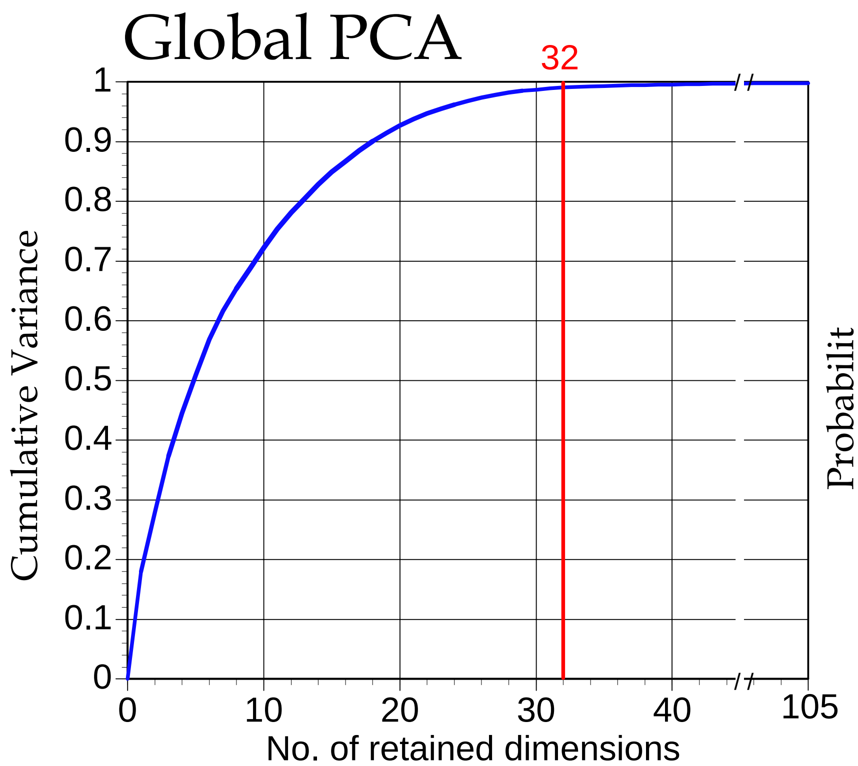
- Subset Selection & Shrinkage
    - Ridge regression, Lasso
- PCA, PCR, PLS
- Comparison of Methods

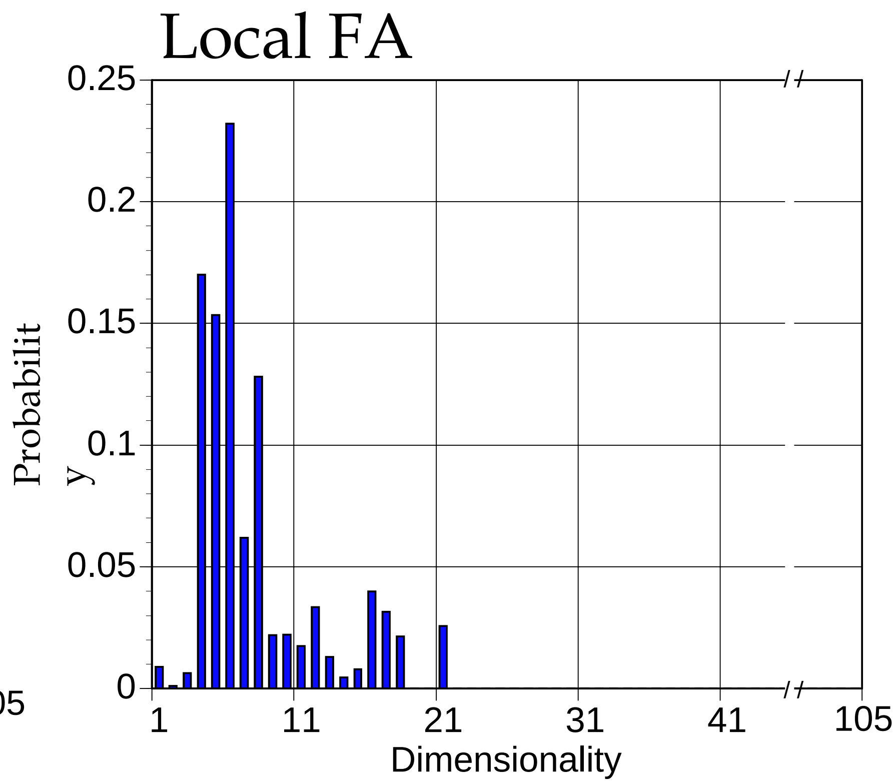
Data From Human Movement

- Measure arm movement and full-body movement of humans and anthropomorphic robots
- Perform local dimensionality analysis with a growing variational mixture of factor analyzers



Dimensionality of Full Body Motion



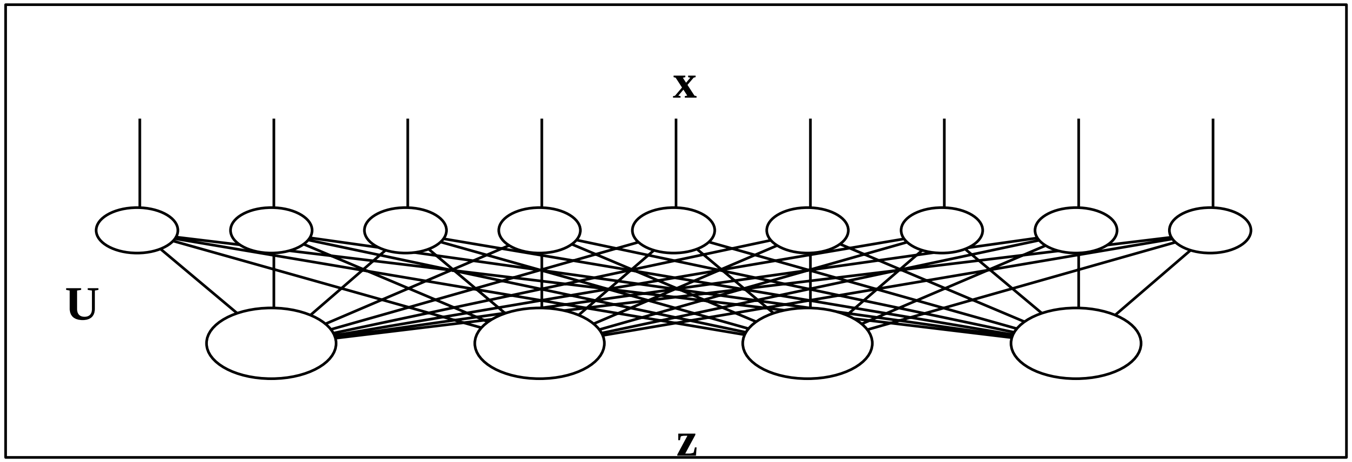


About 8 dimensions in the space formed by joint positions, velocities, and accelerations are needed to model an inverse dynamics model

The Probabilistic Way: Factor Analysis

- Data Generating Model:

$$\mathbf{x} = \mathbf{U}\mathbf{z} + \varepsilon \quad \text{where } \varepsilon = N\left(0, \Omega\right) \text{ and } \mathbf{z} = N\left(0, 1\right)$$



- The parameters U and Ω can be estimated by max. likelihood, in particular the EM algorithm

The EM-Algorithm for Factor Analysis

E – Step :

$$\beta = \mathbf{U}^T \left(\Omega + \mathbf{U}\mathbf{U}^T \right)^{-1}$$

$$E\left\{ \mathbf{z} \mid \mathbf{x} \right\} = \beta \mathbf{x}$$

$$E\left\{ \mathbf{z}\mathbf{z}^T \mid \mathbf{x} \right\} = \mathbf{I} - \beta \mathbf{U} + \beta \mathbf{x}\mathbf{x}^T \beta^T$$

M – Step :

$$\mathbf{U}^{new} = \left(\sum_{n=1}^{N} \mathbf{x}^n E\left(\mathbf{z} \mid \mathbf{x}^n \right)^T \right) \left(\sum_{n=1}^{N} E\left\{ \mathbf{z}\mathbf{z}^T \mid \mathbf{x}^n \right\}^T \right)^{-1}$$

$$\Omega^{new} = \frac{1}{N} diag \left\{ \sum_{n=1}^{N} \mathbf{x}^n \mathbf{x}^{nT} - \mathbf{U}^{new} E\left(\mathbf{z} \mid \mathbf{x}^n \right) \mathbf{x}^{nT} \right\}$$

Factor Analysis for Supervised Learning

- A straightforward extension allows factor analysis to be used for supervised learning:

$$\mathbf{z} = \begin{bmatrix} \mathbf{x} - \bar{\mathbf{x}} \\ \mathbf{y} - \bar{\mathbf{y}} \end{bmatrix}$$

$\mathbf{v} = \tilde{\mathbf{x}}$, i.e., the TRUE (non-noise contaminated inputs

$$\mathbf{U} = [\mathbf{I}, \mathbf{W}]^T$$

- After performing EM on joint data, the network weights are:

$$E\left\{ \begin{bmatrix} y \\ \mathbf{v} \end{bmatrix} \middle| \mathbf{x} \right\} = \begin{bmatrix} \mathbf{W}^T \\ \mathbf{B} \end{bmatrix} \mathbf{x} = \Psi_{21} \Psi_{11}^{-1} \mathbf{x}, \quad \text{where}$$

$$\Psi = \begin{bmatrix} \Omega + \mathbf{U}\mathbf{U}^T & \mathbf{U} \\ \mathbf{U}^T & \mathbf{I} \end{bmatrix} = \begin{bmatrix} \Psi_{11} (= d \times d) & \Psi_{12} (= d \times (c+k)) \\ \Psi_{21} (= (c+k) \times d) & \Psi_{22} (= (c+k) \times (c+k)) \end{bmatrix}$$

d:= dimensionality of observed data

c:= dimensionality of (supervised) outputs

k:= reduced dimensionality

Partial Least Squares (PLS)

- Partial Least Squares is a linear regression methods that includes dimensionality reduction

Build the matrix X and vector y

$$\mathbf{X} = (\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, ..., \tilde{\mathbf{x}}_n)^T$$
$$\mathbf{t} = (t_1, t_2, ..., t_n)^T$$

Projection direction based on input output correlation.

Recursively compute the linear model

$$\text{Initialize}: \mathbf{X}_{\mathbf{res}} = \mathbf{X},\ \mathbf{y} = \mathbf{y}_{\mathbf{res}}$$
$$\text{For } i = 1 : r \ (\#\textit{of projections})$$

Projection direction

$$\mathbf{u}_\mathbf{i} = \mathbf{X}_{\mathbf{res}}^T \mathbf{y}_{\mathbf{res}}$$
$$\mathbf{s} = \mathbf{X}_{\mathbf{res}} \mathbf{u}_\mathbf{i}$$

Projected data

Univariate regressions

$$w_i^{pls} = \frac{\mathbf{s}^T \mathbf{y}_{\mathbf{res}}}{\mathbf{s}^T \mathbf{s}}$$
$$\mathbf{p}_i = \frac{\mathbf{s}^T \mathbf{X}_{\mathbf{res}}}{\mathbf{s}^T \mathbf{s}}$$

residuals

$$\mathbf{y}_{\mathbf{res}} = \mathbf{y}_{\mathbf{res}} - \mathbf{s}\beta_i$$
$$\mathbf{X}_{\mathbf{res}} = \mathbf{X}_{\mathbf{res}} - \mathbf{s}\mathbf{p}_i$$

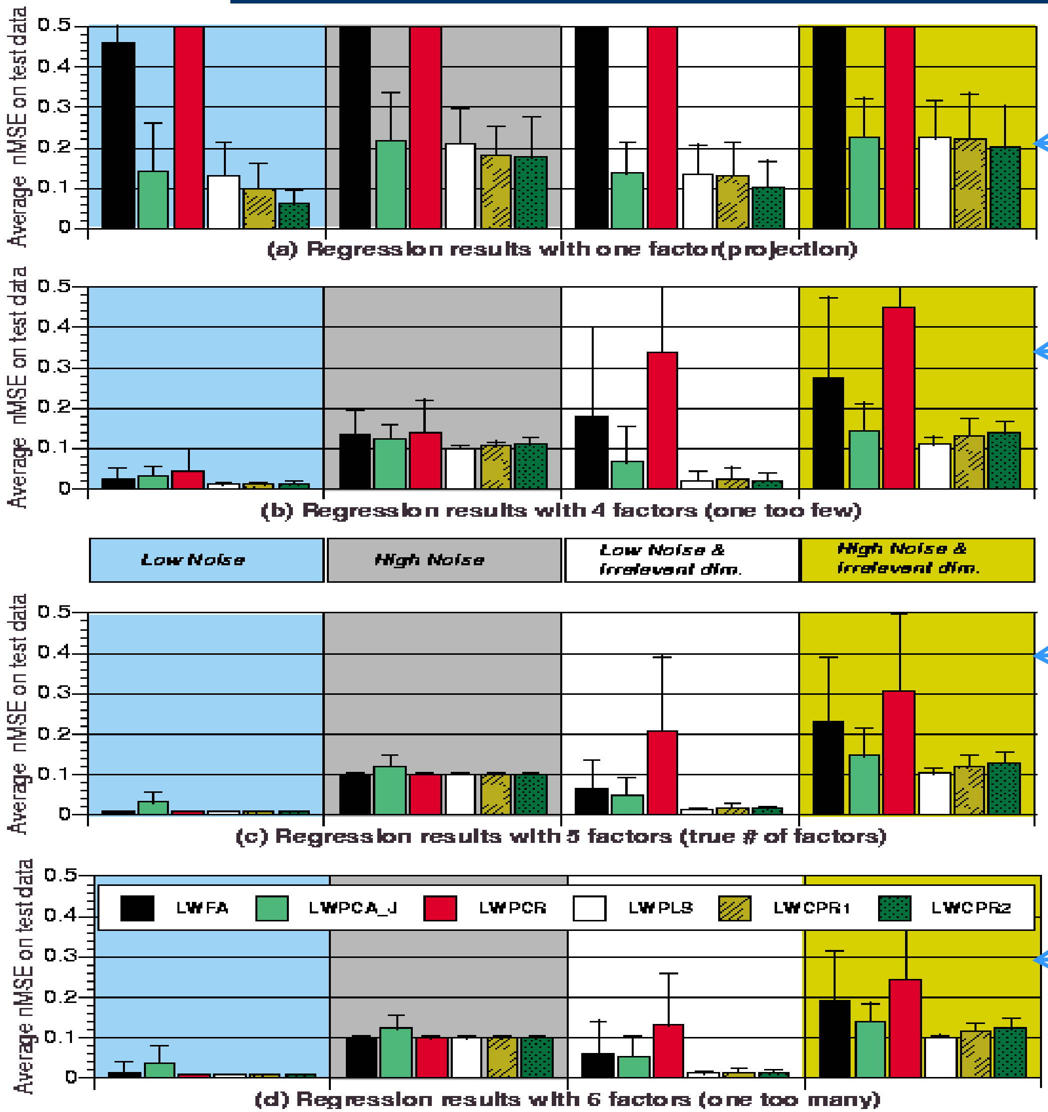
Comparing Dimensionality Red. Methods

FA

PCR

PCA_J

PLS



One Projection

Four Projections

Five Projections

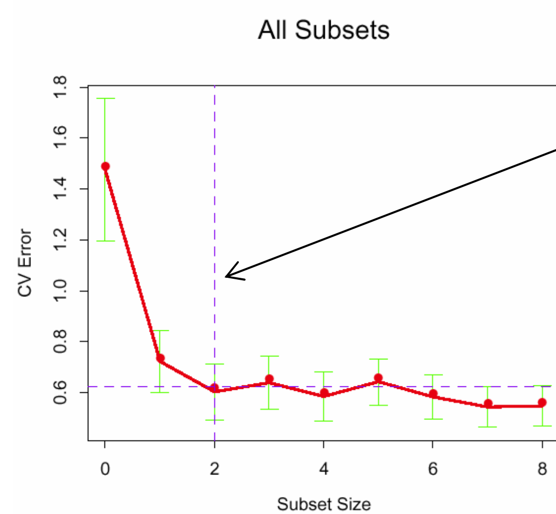
- Correct

Six Projections

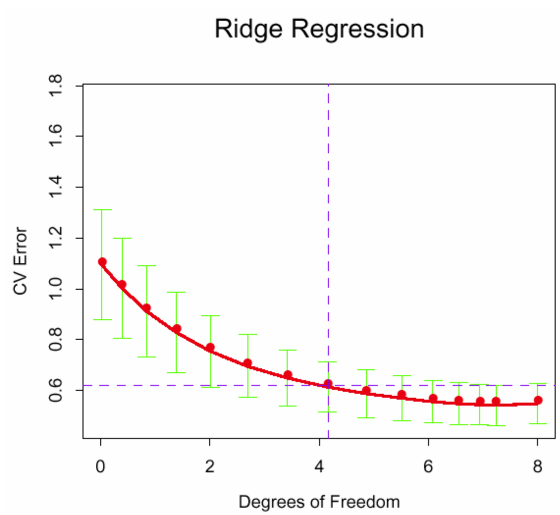
Comparison of Methods (I)

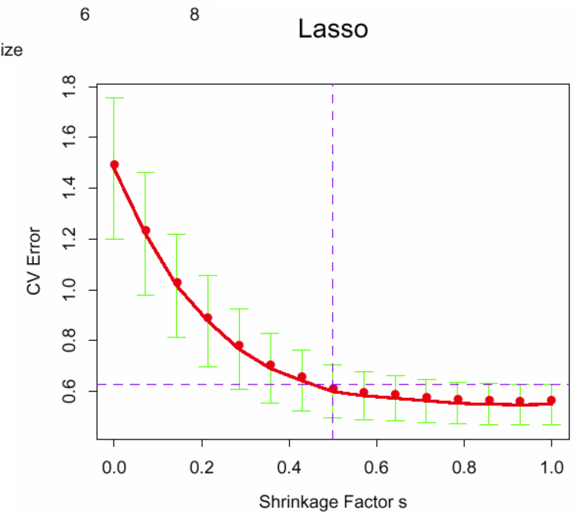
Prostate Cancer Data Example

(pg. 57, Elem. Stat. Analysis)



Least complex model within one S.D. of the best

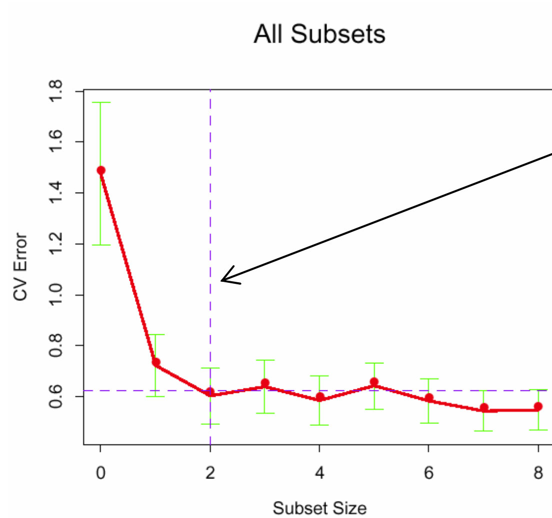




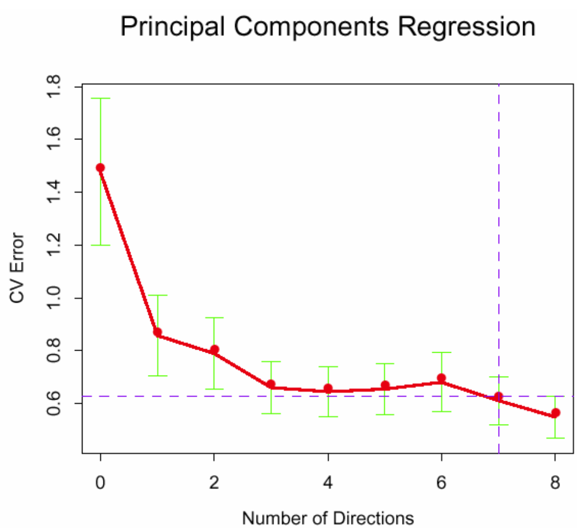
Comparison of Methods (II)

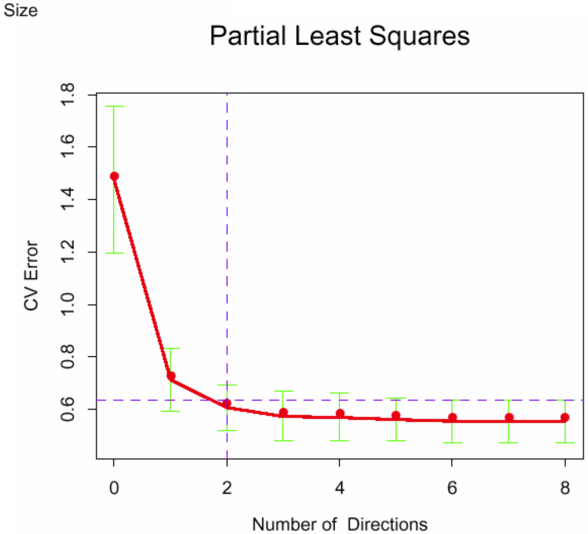
Prostate Cancer Data Example

(pg. 57, Elem. Stat. Analysis)



Least complex model within one S.D. of the best

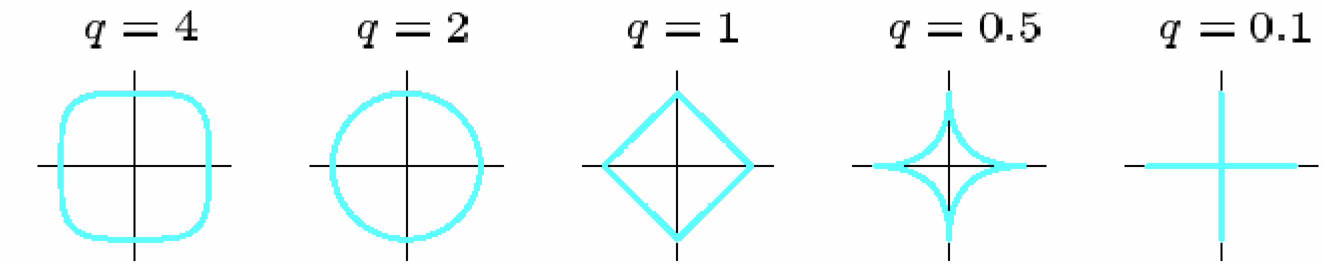




Generalization of Shrinkage Methods

$$\hat{w}^{gen} = \arg\min_{w} \left\{ \sum_{i=1}^{N} (t_i - w_0 - \sum_{j=1}^{M} x_{ij} w_j)^2 + \lambda \sum_{j=1}^{M} \left| w_j \right|^q \right\} \quad \textit{for} \;\; q \geq 0.$$

$q = 0 : \textit{variable subset selection}$
$q = 1 : \textit{lasso}$
$q = 2 : \textit{ridge regression}$



Contours of constant value of $\sum_{j=1}^{M} \left| w_j \right|^q$ for given value of q.

Error- Regularization Tradeoff