

# **Regression**

## **ML - Supervised Learning**

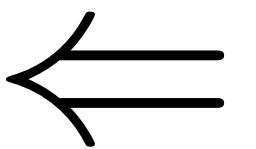
**Dan Calderone - Win22**

# Data

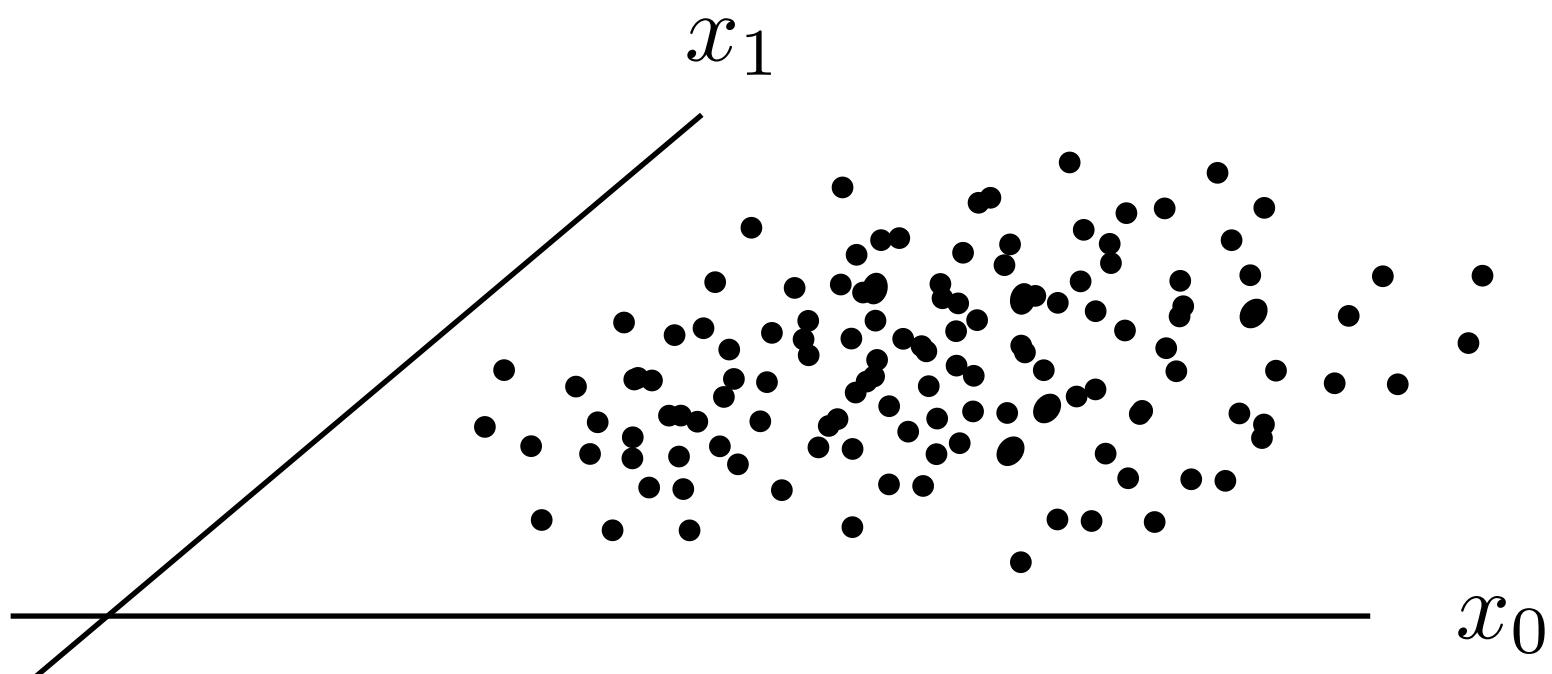
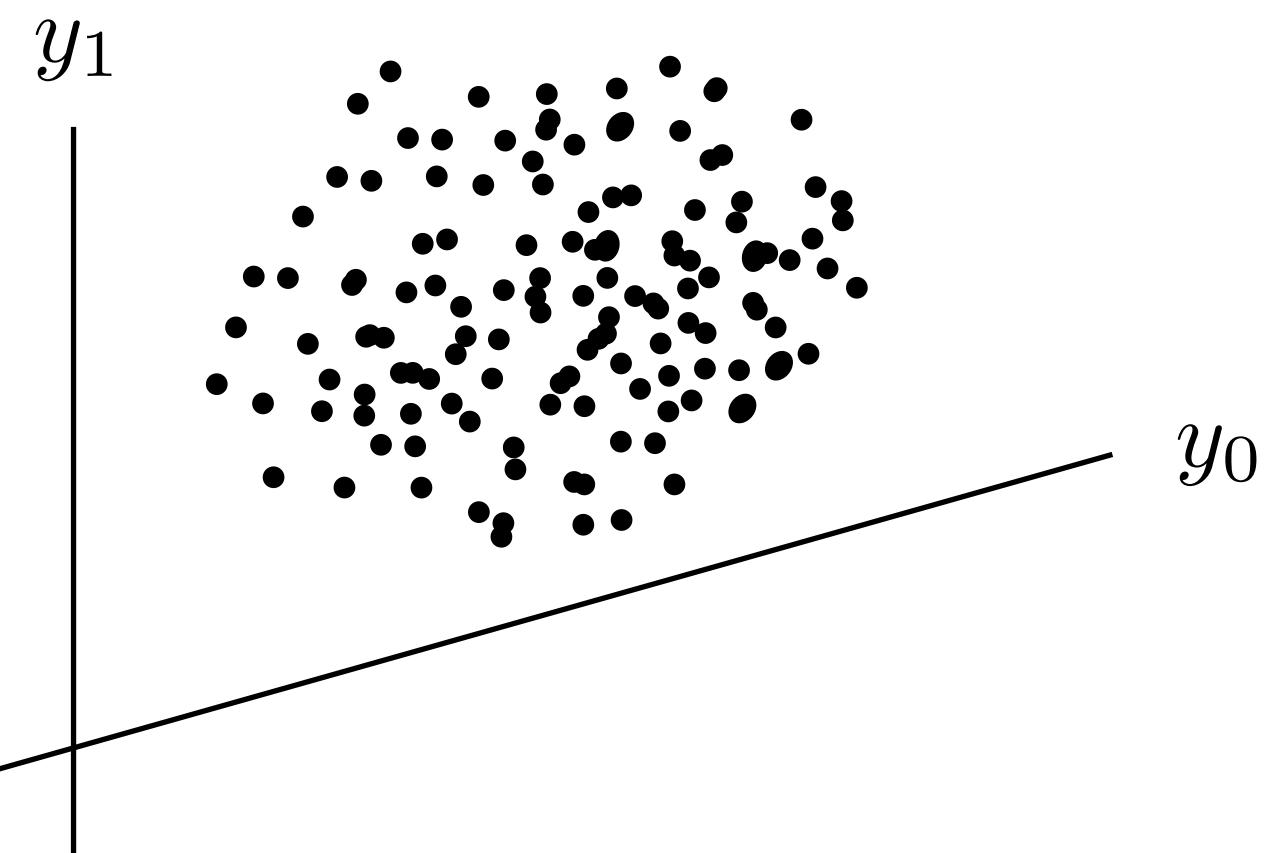
## OUTPUTS

### (Dependent Variables)

$y_{00}$	$\dots$	$y_{0m}$	$\gamma_{00}$	$\dots$	$\gamma_{0m'}$
$y_{10}$	$\dots$	$y_{1m}$	$\gamma_{10}$	$\dots$	$\gamma_{1m'}$
$y_{20}$	$\dots$	$y_{2m}$	$\gamma_{20}$	$\dots$	$\gamma_{2m'}$
$y_{30}$	$\dots$	$y_{3m}$	$\gamma_{30}$	$\dots$	$\gamma_{3m'}$
$y_{40}$	$\dots$	$y_{4m}$	$\gamma_{40}$	$\dots$	$\gamma_{4m'}$
$\vdots$		$\vdots$	$\vdots$		$\vdots$
$y_{T0}$	$\dots$	$y_{Tm}$	$\gamma_{T0}$	$\dots$	$\gamma_{Tm'}$



$x_{00}$	$\dots$	$x_{0n}$	$\xi_{00}$	$\dots$	$\xi_{0n'}$
$x_{10}$	$\dots$	$x_{1n}$	$\xi_{10}$	$\dots$	$\xi_{1n'}$
$x_{20}$	$\dots$	$x_{2n}$	$\xi_{20}$	$\dots$	$\xi_{2n'}$
$x_{30}$	$\dots$	$x_{3n}$	$\xi_{30}$	$\dots$	$\xi_{3n'}$
$x_{40}$	$\dots$	$x_{4n}$	$\xi_{40}$	$\dots$	$\xi_{4n'}$
$\vdots$		$\vdots$	$\vdots$		$\vdots$
$x_{T0}$	$\dots$	$x_{Tn}$	$\xi_{T0}$	$\dots$	$\xi_{Tn'}$

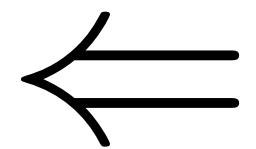


# Data

## OUTPUTS

### (Dependent Variables)

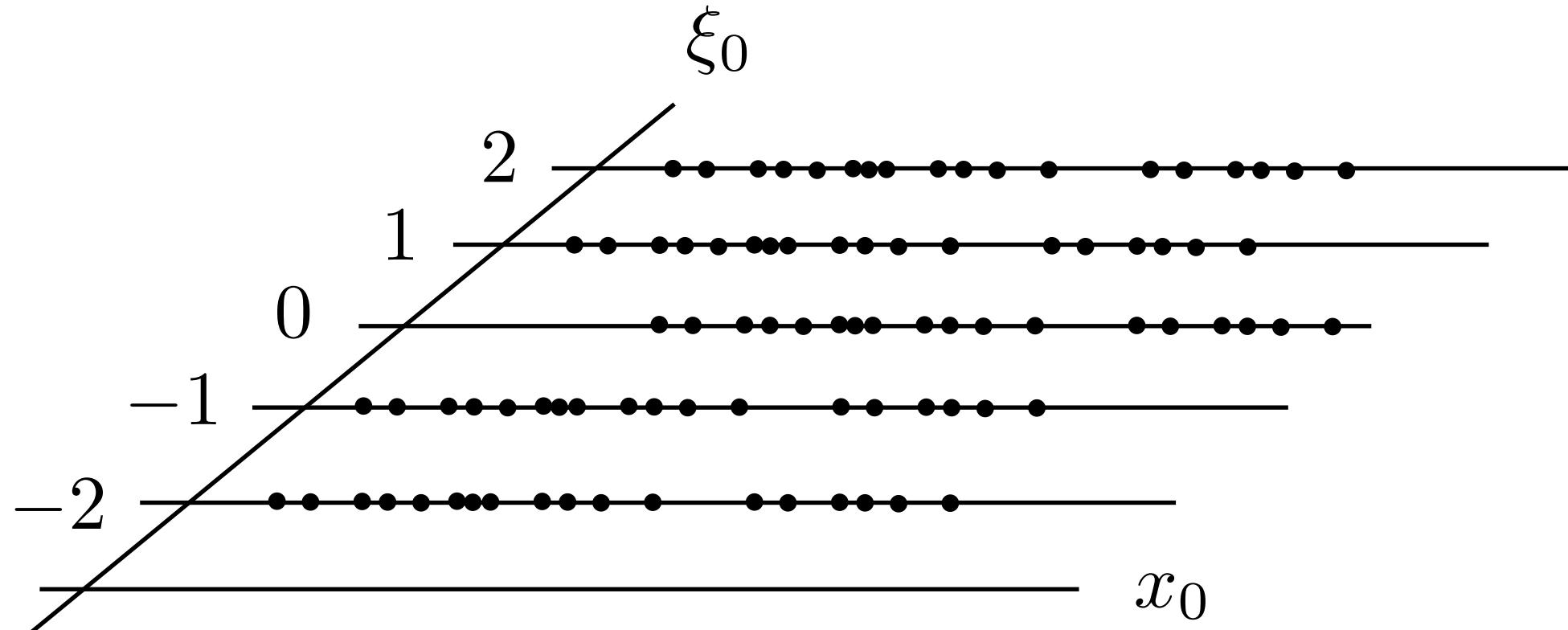
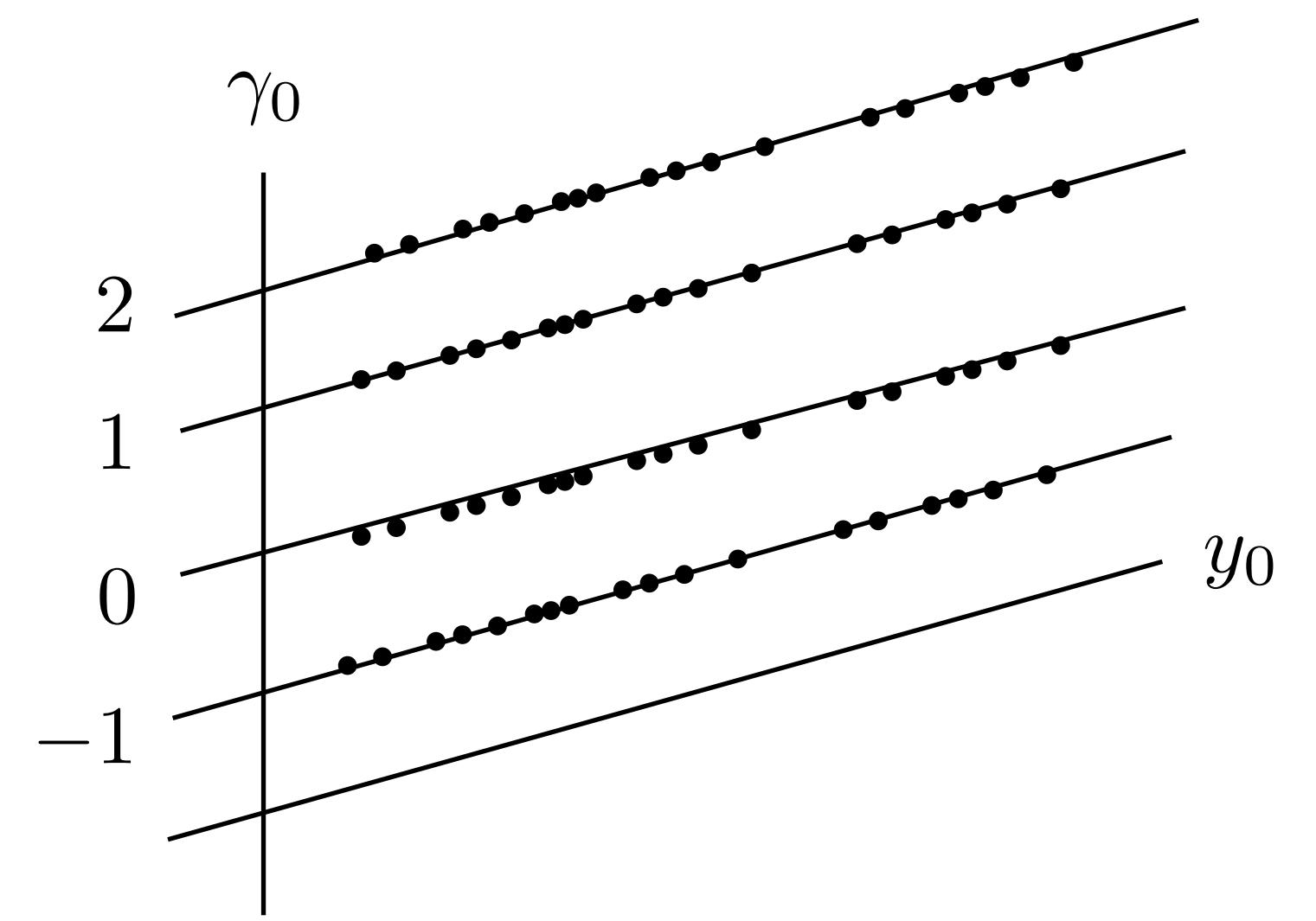
$$\begin{bmatrix} y_{00} & \cdots & y_{0m} & \boxed{y_{00}} & \cdots & y_{0m'} \\ y_{10} & \cdots & y_{1m} & \boxed{\gamma_{10}} & \cdots & \gamma_{1m'} \\ y_{20} & \cdots & y_{2m} & \boxed{\gamma_{20}} & \cdots & \gamma_{2m'} \\ y_{30} & \cdots & y_{3m} & \boxed{\gamma_{30}} & \cdots & \gamma_{3m'} \\ y_{40} & \cdots & y_{4m} & \boxed{\gamma_{40}} & \cdots & \gamma_{4m'} \\ \vdots & & \vdots & \vdots & & \vdots \\ y_{T0} & \cdots & y_{Tm} & \boxed{\gamma_{T0}} & \cdots & \gamma_{Tm'} \end{bmatrix}$$



## INPUTS

### (Independent Variables)

$$\begin{bmatrix} x_{00} & \cdots & x_{0n} & \boxed{x_{00}} & \cdots & x_{0n'} \\ x_{10} & \cdots & x_{1n} & \boxed{\xi_{10}} & \cdots & \xi_{1n'} \\ x_{20} & \cdots & x_{2n} & \boxed{\xi_{20}} & \cdots & \xi_{2n'} \\ x_{30} & \cdots & x_{3n} & \boxed{\xi_{30}} & \cdots & \xi_{3n'} \\ x_{40} & \cdots & x_{4n} & \boxed{\xi_{40}} & \cdots & \xi_{4n'} \\ \vdots & & \vdots & \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} & \boxed{\xi_{T0}} & \cdots & \xi_{Tn'} \end{bmatrix}$$

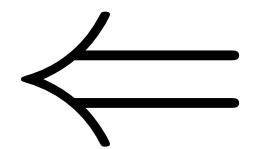


# Data

## OUTPUTS

(Dependent Variables)

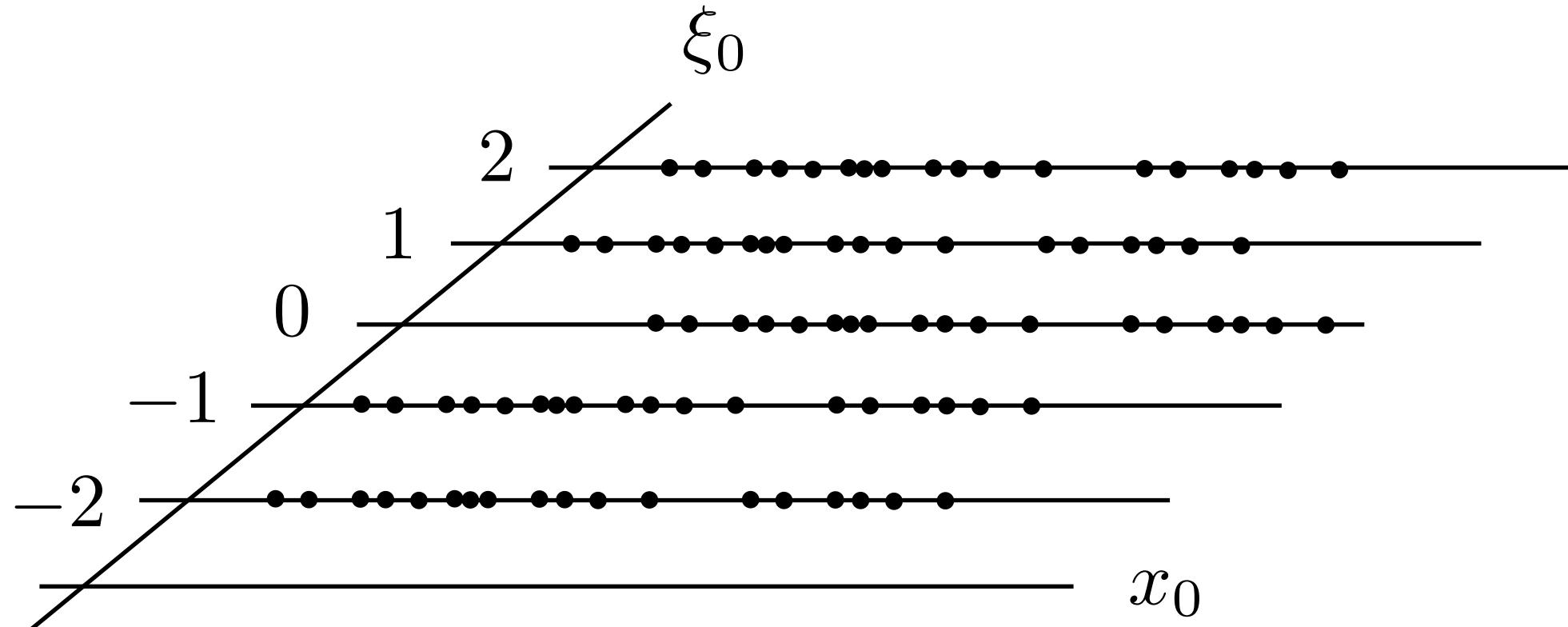
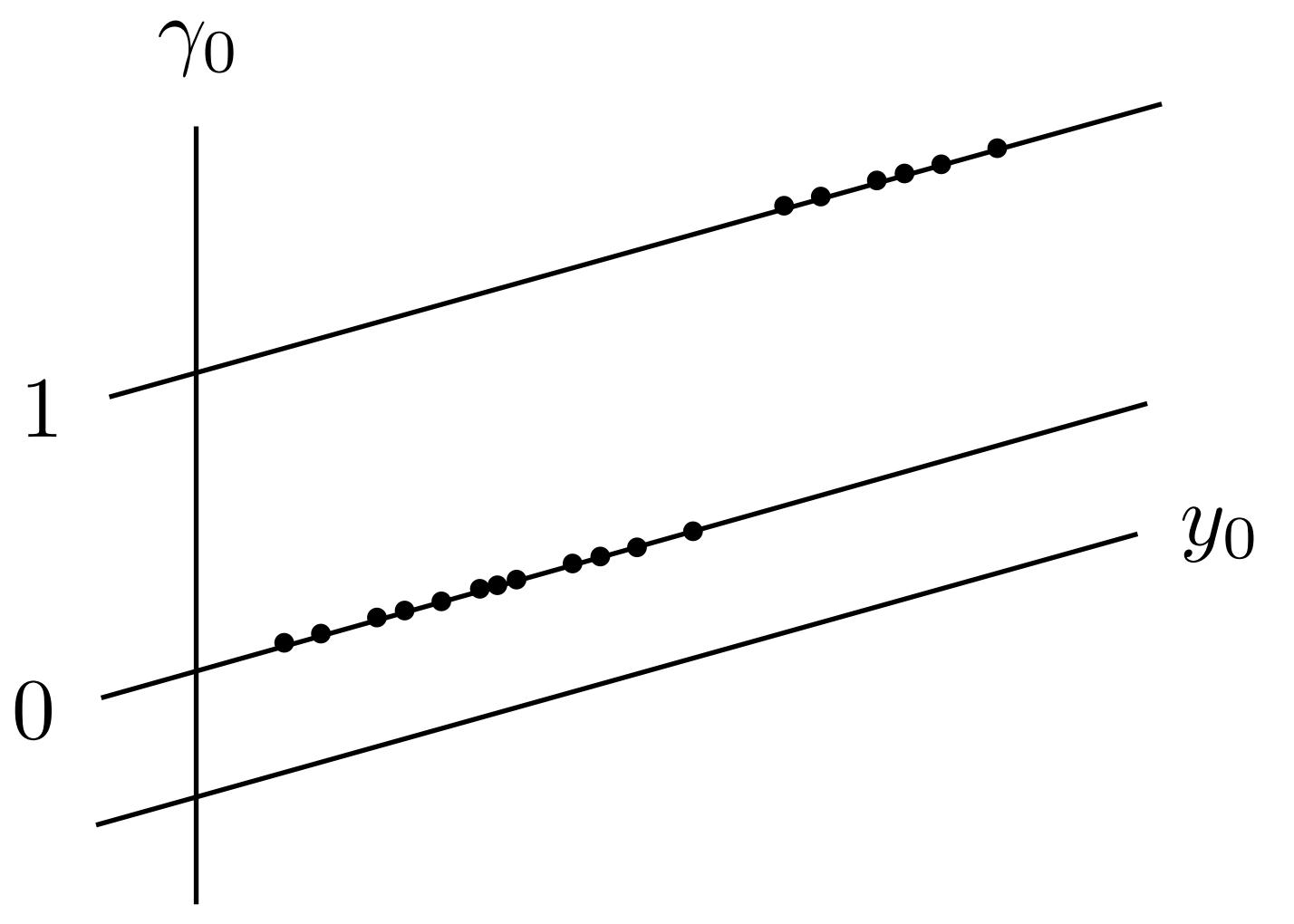
$$\begin{bmatrix} y_{00} & \cdots & y_{0m} & \boxed{y_{00}} & \cdots & y_{0m'} \\ y_{10} & \cdots & y_{1m} & \boxed{\gamma_{10}} & \cdots & \gamma_{1m'} \\ y_{20} & \cdots & y_{2m} & \boxed{\gamma_{20}} & \cdots & \gamma_{2m'} \\ y_{30} & \cdots & y_{3m} & \boxed{\gamma_{30}} & \cdots & \gamma_{3m'} \\ y_{40} & \cdots & y_{4m} & \boxed{\gamma_{40}} & \cdots & \gamma_{4m'} \\ \vdots & & \vdots & \vdots & & \vdots \\ y_{T0} & \cdots & y_{Tm} & \boxed{\gamma_{T0}} & \cdots & \gamma_{Tm'} \end{bmatrix}$$



## INPUTS

(Independent Variables)

$$\begin{bmatrix} x_{00} & \cdots & x_{0n} & \boxed{x_{00}} & \cdots & x_{0n'} \\ x_{10} & \cdots & x_{1n} & \boxed{\xi_{10}} & \cdots & \xi_{1n'} \\ x_{20} & \cdots & x_{2n} & \boxed{\xi_{20}} & \cdots & \xi_{2n'} \\ x_{30} & \cdots & x_{3n} & \boxed{\xi_{30}} & \cdots & \xi_{3n'} \\ x_{40} & \cdots & x_{4n} & \boxed{\xi_{40}} & \cdots & \xi_{4n'} \\ \vdots & & \vdots & \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} & \boxed{\xi_{T0}} & \cdots & \xi_{Tn'} \end{bmatrix}$$



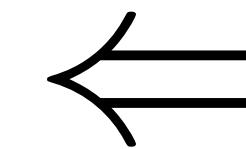
# Functions

OUTPUTS  
(Dependent Variables)

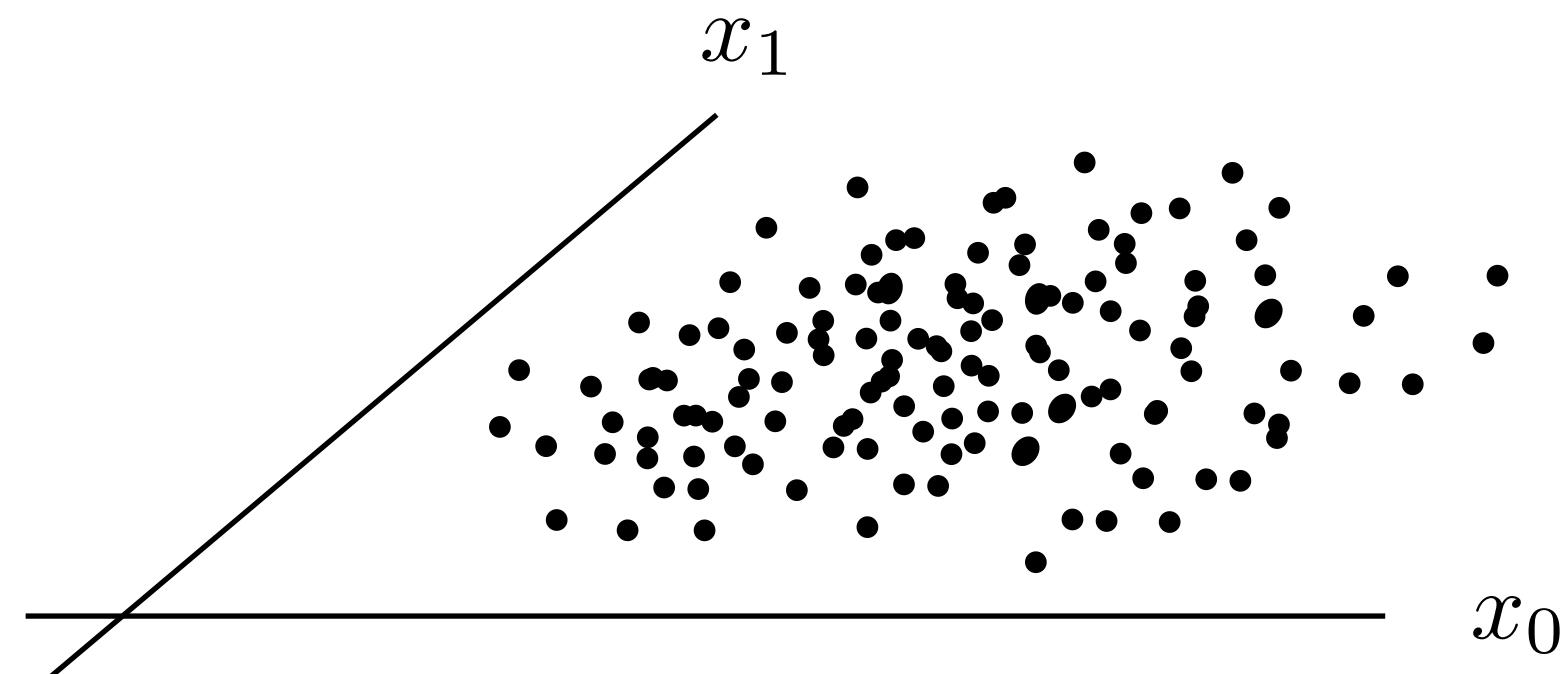
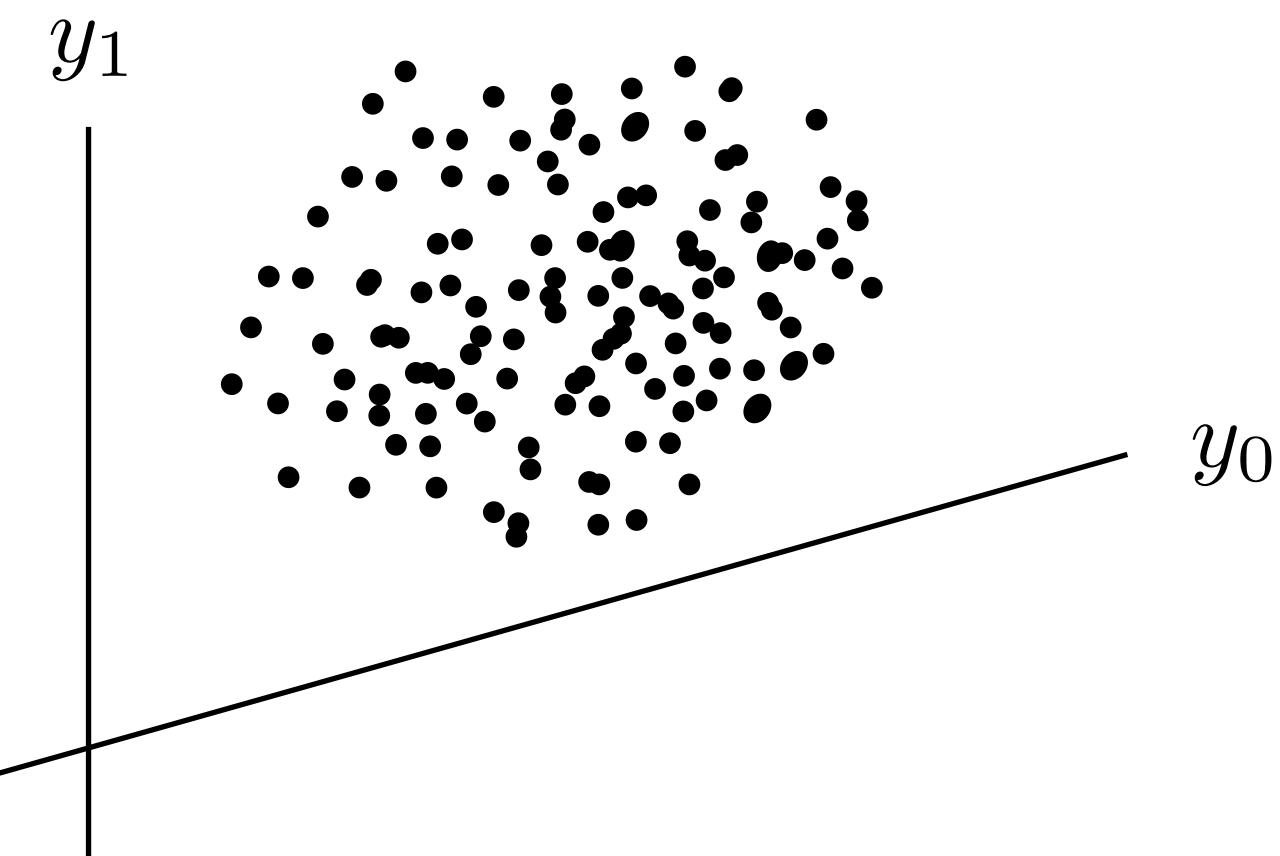
$$y_t = f(x_t)$$

INPUTS  
(Independent Variables)

$y_{00}$	$\cdots$	$y_{0m}$	$\gamma_{00}$	$\cdots$	$\gamma_{0m'}$
$y_{10}$	$\cdots$	$y_{1m}$	$\gamma_{10}$	$\cdots$	$\gamma_{1m'}$
$y_{20}$	$\cdots$	$y_{2m}$	$\gamma_{20}$	$\cdots$	$\gamma_{2m'}$
$y_{30}$	$\cdots$	$y_{3m}$	$\gamma_{30}$	$\cdots$	$\gamma_{3m'}$
$y_{40}$	$\cdots$	$y_{4m}$	$\gamma_{40}$	$\cdots$	$\gamma_{4m'}$
$\vdots$		$\vdots$	$\vdots$		$\vdots$
$y_{T0}$	$\cdots$	$y_{Tm}$	$\gamma_{T0}$	$\cdots$	$\gamma_{Tm'}$



$f$	$[$	$($	$x_{00}$	$\cdots$	$x_{0n}$	$\xi_{00}$	$\cdots$	$\xi_{0n'}$	$)$	$]$
			$x_{10}$	$\cdots$	$x_{1n}$	$\xi_{10}$	$\cdots$	$\xi_{1n'}$		
			$x_{20}$	$\cdots$	$x_{2n}$	$\xi_{20}$	$\cdots$	$\xi_{2n'}$		
			$x_{30}$	$\cdots$	$x_{3n}$	$\xi_{30}$	$\cdots$	$\xi_{3n'}$		
			$x_{40}$	$\cdots$	$x_{4n}$	$\xi_{40}$	$\cdots$	$\xi_{4n'}$		
			$\vdots$		$\vdots$	$\vdots$		$\vdots$		$\vdots$
			$x_{T0}$	$\cdots$	$x_{Tn}$	$\xi_{T0}$	$\cdots$	$\xi_{Tn'}$		



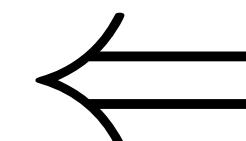
# Functions

OUTPUTS  
(Dependent Variables)

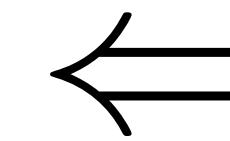
$$y_t = f(x_t)$$

INPUTS  
(Independent Variables)

$$\begin{bmatrix} y_{00} & \cdots & y_{0m} \\ y_{10} & \cdots & y_{1m} \\ y_{20} & \cdots & y_{2m} \\ y_{30} & \cdots & y_{3m} \\ y_{40} & \cdots & y_{4m} \\ \vdots & & \vdots \\ y_{T0} & \cdots & y_{Tm} \end{bmatrix}$$

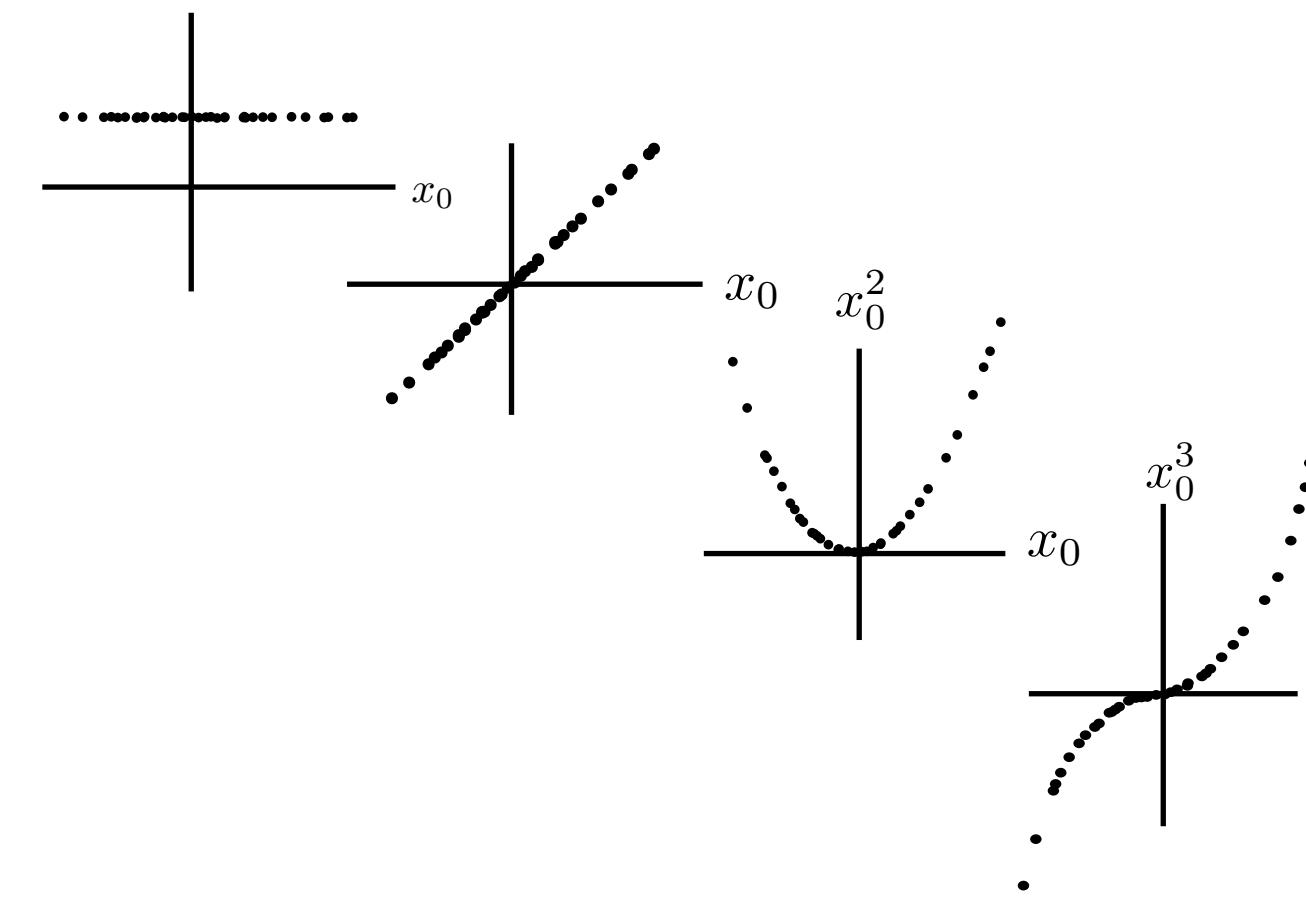
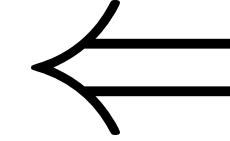
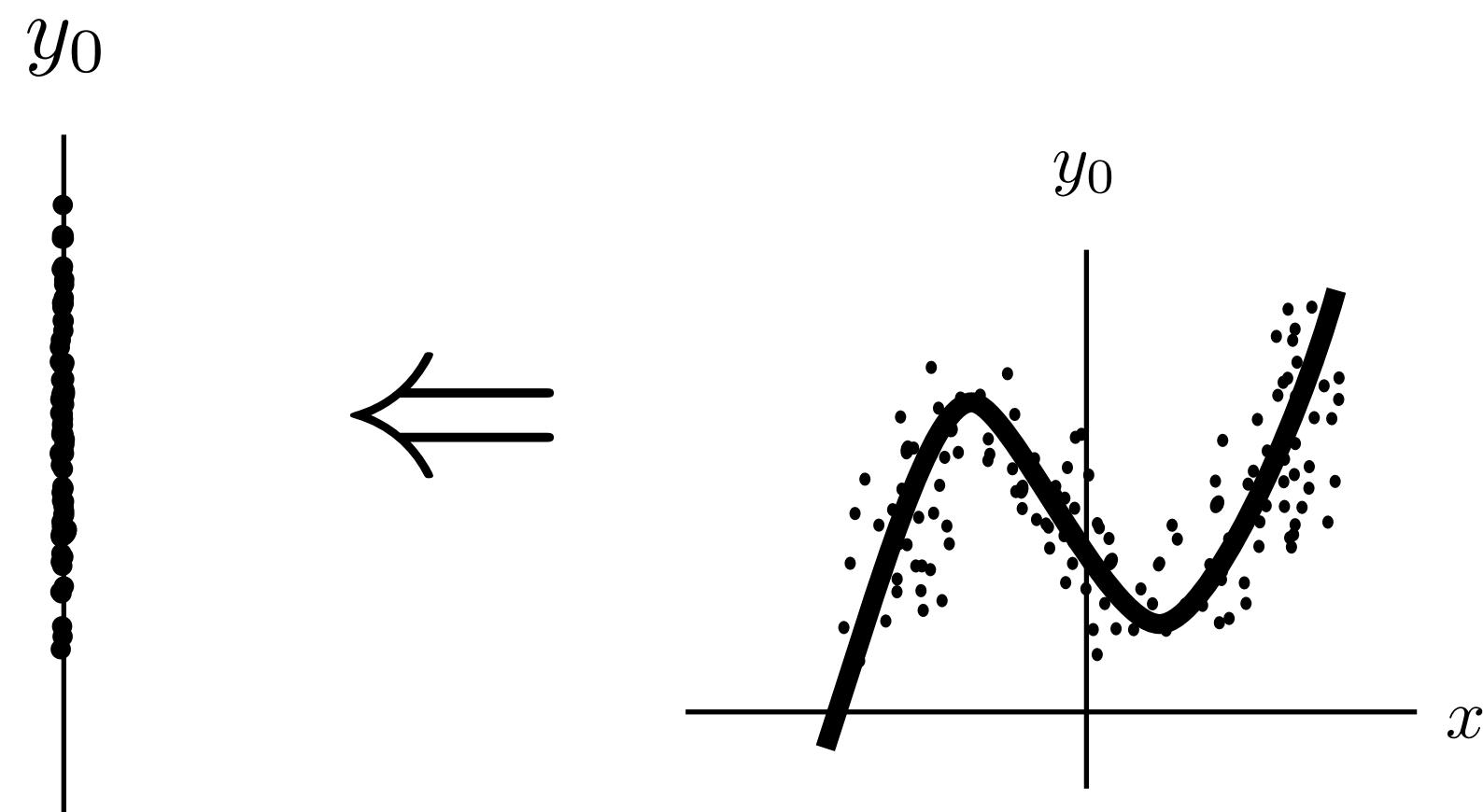


$$\begin{bmatrix} h_{00}(x_0, \xi_0) & \cdots & h_{0n}(x_0, \xi_0) \\ h_{10}(x_1, \xi_1) & \cdots & h_{1n}(x_1, \xi_1) \\ h_{20}(x_2, \xi_2) & \cdots & h_{2n}(x_2, \xi_2) \\ h_{30}(x_3, \xi_3) & \cdots & h_{3n}(x_3, \xi_3) \\ h_{40}(x_4, \xi_4) & \cdots & h_{4n}(x_4, \xi_4) \\ \vdots & & \vdots \\ h_{T0}(x_T, \xi_T) & \cdots & h_{Tn}(x_T, \xi_T) \end{bmatrix}$$



$$h_t \begin{bmatrix} (x_{00} & \cdots & x_{0n} & \xi_{00} & \cdots & \xi_{0n'}) \\ x_{10} & \cdots & x_{1n} & \xi_{10} & \cdots & \xi_{1n'} \\ x_{20} & \cdots & x_{2n} & \xi_{20} & \cdots & \xi_{2n'} \\ x_{30} & \cdots & x_{3n} & \xi_{30} & \cdots & \xi_{3n'} \\ x_{40} & \cdots & x_{4n} & \xi_{40} & \cdots & \xi_{4n'} \\ \vdots & & \vdots & \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} & \xi_{T0} & \cdots & \xi_{Tn'} \end{bmatrix}$$

with BASIS FUNCTIONS



$$h_t(x_t) = [1 \ x_{t0} \ x_{t0}^2 \ x_{t0}^3]$$

# Functions

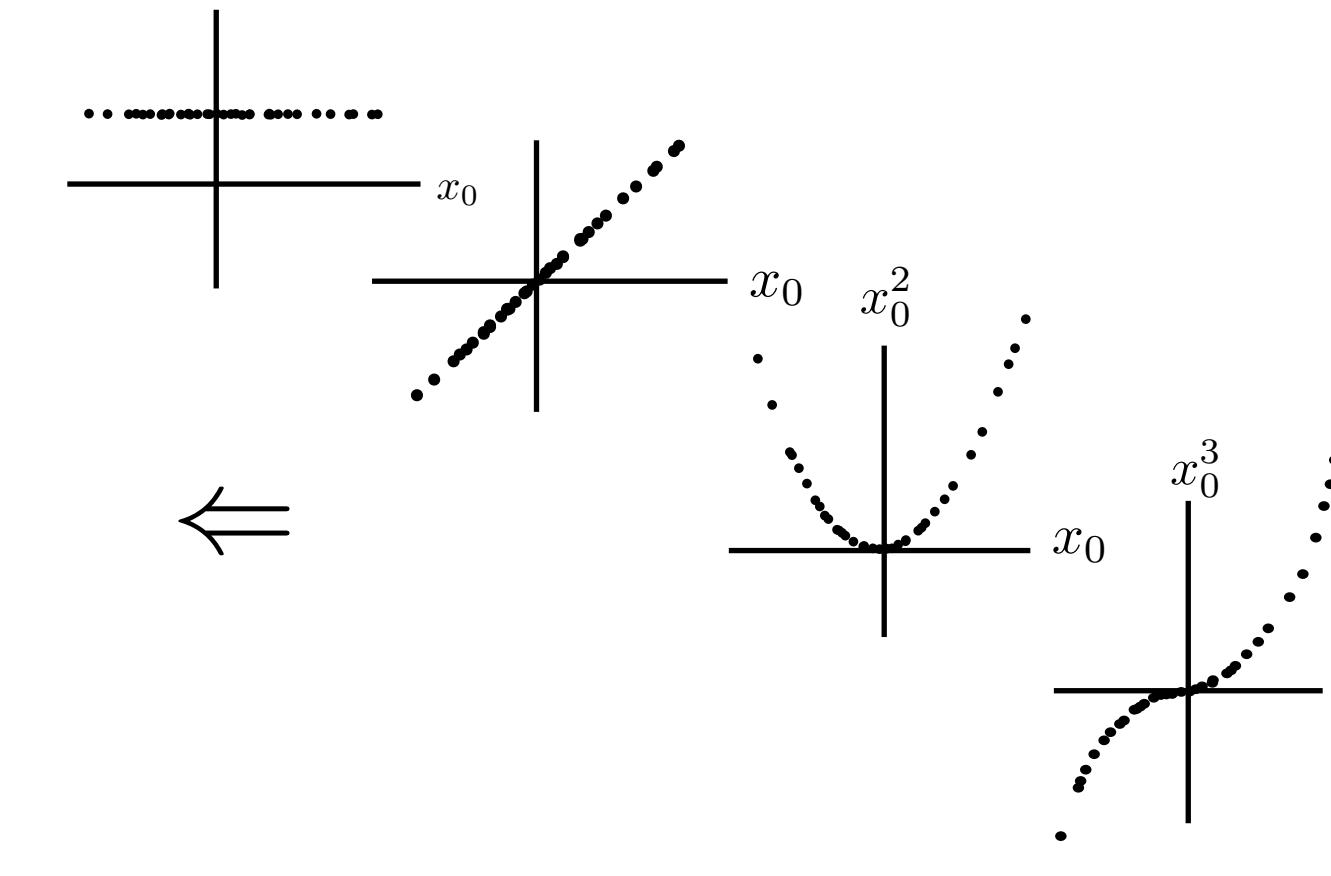
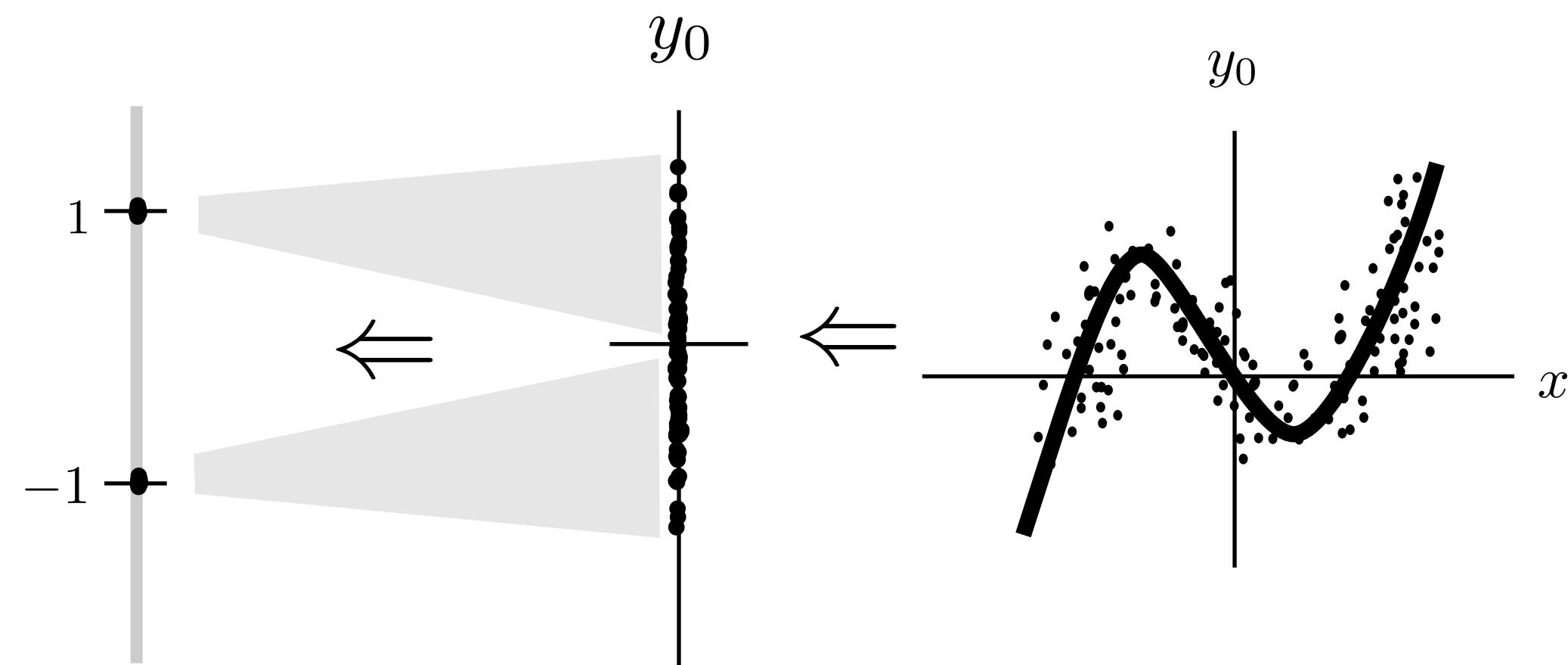
OUTPUTS  
(Dependent Variables)

$$y_t = f(x_t)$$

INPUTS  
(Independent Variables)

$$\begin{bmatrix} \gamma_{00} \\ \gamma_{10} \\ \gamma_{20} \\ \gamma_{30} \\ \gamma_{40} \\ \vdots \\ \gamma_{T0} \end{bmatrix} \leftarrow \begin{bmatrix} y_{00} \\ y_{10} \\ y_{20} \\ y_{30} \\ y_{40} \\ \vdots \\ y_{T0} \end{bmatrix} \leftarrow \begin{bmatrix} h_{00}(x_0, \xi_0) & \cdots & h_{0n}(x_0, \xi_0) \\ h_{10}(x_1, \xi_1) & \cdots & h_{1n}(x_1, \xi_1) \\ h_{20}(x_2, \xi_2) & \cdots & h_{2n}(x_2, \xi_2) \\ h_{30}(x_3, \xi_3) & \cdots & h_{3n}(x_3, \xi_3) \\ h_{40}(x_4, \xi_4) & \cdots & h_{4n}(x_4, \xi_4) \\ \vdots & & \vdots \\ h_{T0}(x_T, \xi_T) & \cdots & h_{Tn}(x_T, \xi_T) \end{bmatrix} \leftarrow h_t \begin{bmatrix} (x_{00} & \cdots & x_{0n} & \xi_{00} & \cdots & \xi_{0n'}) \\ x_{10} & \cdots & x_{1n} & \xi_{10} & \cdots & \xi_{1n'} \\ x_{20} & \cdots & x_{2n} & \xi_{20} & \cdots & \xi_{2n'} \\ x_{30} & \cdots & x_{3n} & \xi_{30} & \cdots & \xi_{3n'} \\ x_{40} & \cdots & x_{4n} & \xi_{40} & \cdots & \xi_{4n'} \\ \vdots & & \vdots & \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} & \xi_{T0} & \cdots & \xi_{Tn'} \end{bmatrix}$$

with BASIS FUNCTIONS



$$h_t(x_t) = [1 \ x_{t0} \ x_{t0}^2 \ x_{t0}^3]$$

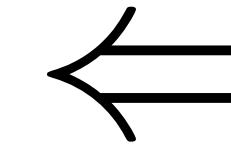
# Linear Regression

OUTPUTS  
(Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS  
(Independent Variables)

$$\begin{bmatrix} y_{00} & \cdots & y_{0m} & \gamma_{00} & \cdots & \gamma_{0m'} \\ y_{10} & \cdots & y_{1m} & \gamma_{10} & \cdots & \gamma_{1m'} \\ y_{20} & \cdots & y_{2m} & \gamma_{20} & \cdots & \gamma_{2m'} \\ y_{30} & \cdots & y_{3m} & \gamma_{30} & \cdots & \gamma_{3m'} \\ y_{40} & \cdots & y_{4m} & \gamma_{40} & \cdots & \gamma_{4m'} \\ \vdots & & \vdots & \vdots & & \vdots \\ y_{T0} & \cdots & y_{Tm} & \gamma_{T0} & \cdots & \gamma_{Tm'} \end{bmatrix}$$



$$\begin{bmatrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \\ \vdots \\ \xi_{T0} \end{bmatrix} \cdots \begin{bmatrix} \xi_{0n'} \\ \xi_{1n'} \\ \vdots \\ \xi_{2n'} \\ \vdots \\ \xi_{3n'} \\ \vdots \\ \xi_{4n'} \\ \vdots \\ \vdots \\ \xi_{Tn'} \end{bmatrix}$$

$$y = X\theta$$

$y_0$

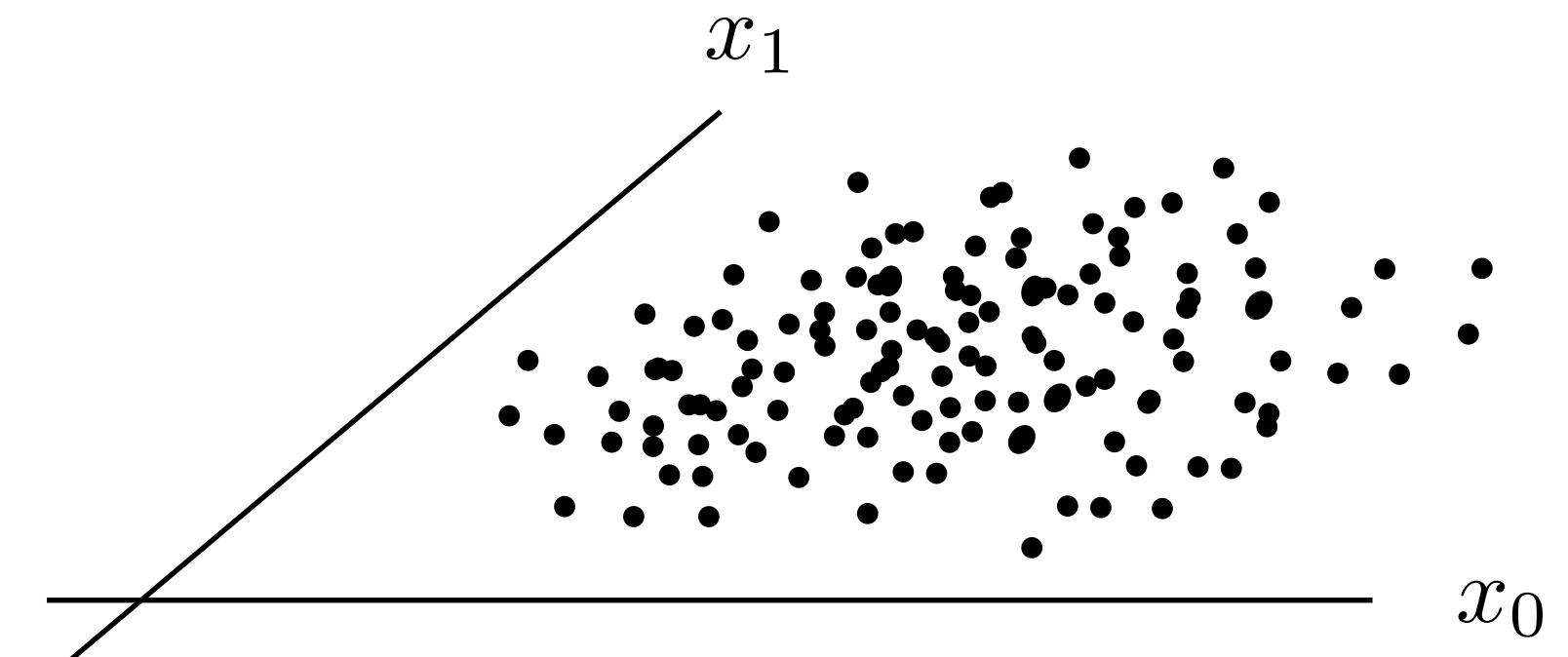


COST:

$$\min ||y - X\theta||^2$$

SOLN:

$$\theta = (X^T X)^{-1} X^T y$$



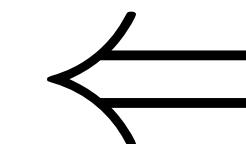
# Linear Regression

OUTPUTS  
(Dependent Variables)

$$y_t = \theta^T x_t$$

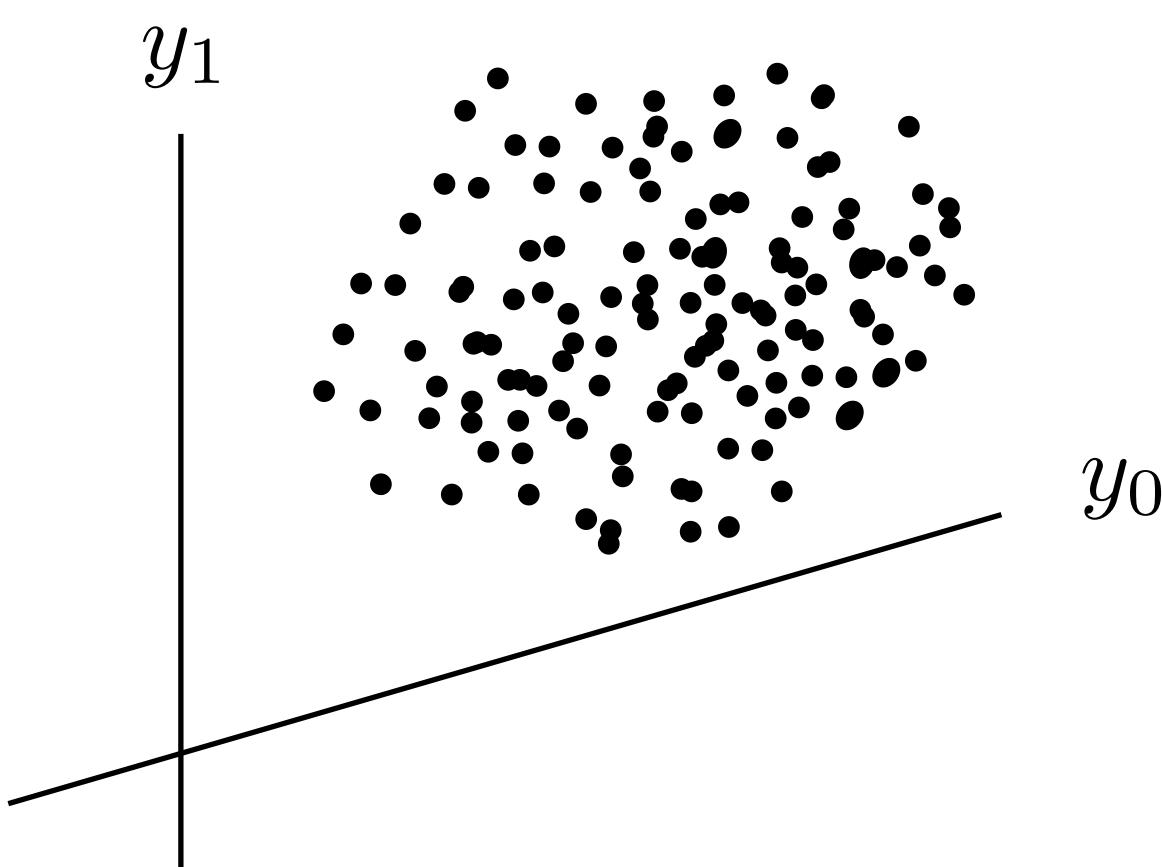
INPUTS  
(Independent Variables)

$$\begin{bmatrix} y_{00} & \cdots & y_{0m} \\ y_{10} & \cdots & y_{1m} \\ y_{20} & \cdots & y_{2m} \\ y_{30} & \cdots & y_{3m} \\ y_{40} & \cdots & y_{4m} \\ \vdots & & \vdots \\ y_{T0} & \cdots & y_{Tm} \end{bmatrix} \quad \begin{bmatrix} \gamma_{00} & \cdots & \gamma_{0m'} \\ \gamma_{10} & \cdots & \gamma_{1m'} \\ \gamma_{20} & \cdots & \gamma_{2m'} \\ \gamma_{30} & \cdots & \gamma_{3m'} \\ \gamma_{40} & \cdots & \gamma_{4m'} \\ \vdots & & \vdots \\ \gamma_{T0} & \cdots & \gamma_{Tm'} \end{bmatrix}$$



$$\begin{bmatrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{bmatrix} \quad \begin{bmatrix} \theta_{00} & \cdots & \theta_{0m'} \\ \xi_{10} & \cdots & \xi_{1n} \\ \xi_{20} & \cdots & \xi_{2n} \\ \theta_{n0} & \cdots & \theta_{nm'} \\ \xi_{40} & \cdots & \xi_{4n'} \\ \vdots & & \vdots \\ \xi_{T0} & \cdots & \xi_{Tn'} \end{bmatrix}$$

$$y = X\theta$$

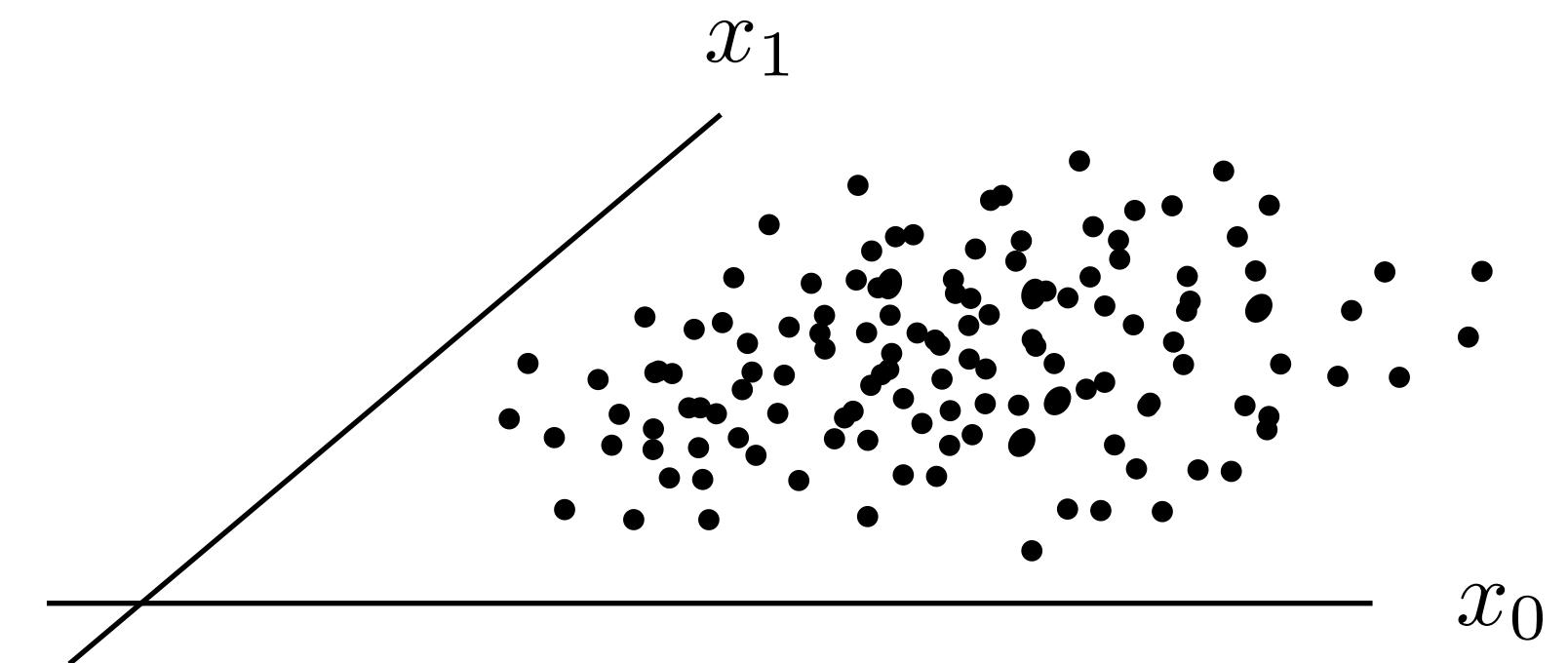


COST:

$$\min ||y - X\theta||^2$$

SOLN:

$$\theta = (X^T X)^{-1} X^T y$$



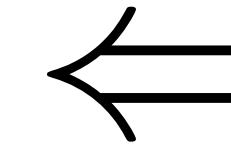
# Linear Regression

OUTPUTS  
(Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS  
(Independent Variables)

$$\begin{bmatrix} y_{00} & \cdots & y_{0m} & \gamma_{00} & \cdots & \gamma_{0m'} \\ y_{10} & \cdots & y_{1m} & \gamma_{10} & \cdots & \gamma_{1m'} \\ y_{20} & \cdots & y_{2m} & \gamma_{20} & \cdots & \gamma_{2m'} \\ y_{30} & \cdots & y_{3m} & \gamma_{30} & \cdots & \gamma_{3m'} \\ y_{40} & \cdots & y_{4m} & \gamma_{40} & \cdots & \gamma_{4m'} \\ \vdots & & \vdots & \vdots & & \vdots \\ y_{T0} & \cdots & y_{Tm} & \gamma_{T0} & \cdots & \gamma_{Tm'} \end{bmatrix}$$



$$\begin{bmatrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \\ \vdots \\ \xi_{T0} \end{bmatrix} \cdots \begin{bmatrix} \xi_{0n'} \\ \xi_{1n'} \\ \vdots \\ \xi_{2n'} \\ \vdots \\ \xi_{3n'} \\ \vdots \\ \xi_{4n'} \\ \vdots \\ \vdots \\ \xi_{Tn'} \end{bmatrix}$$

$$y = X\theta$$

$y_0$

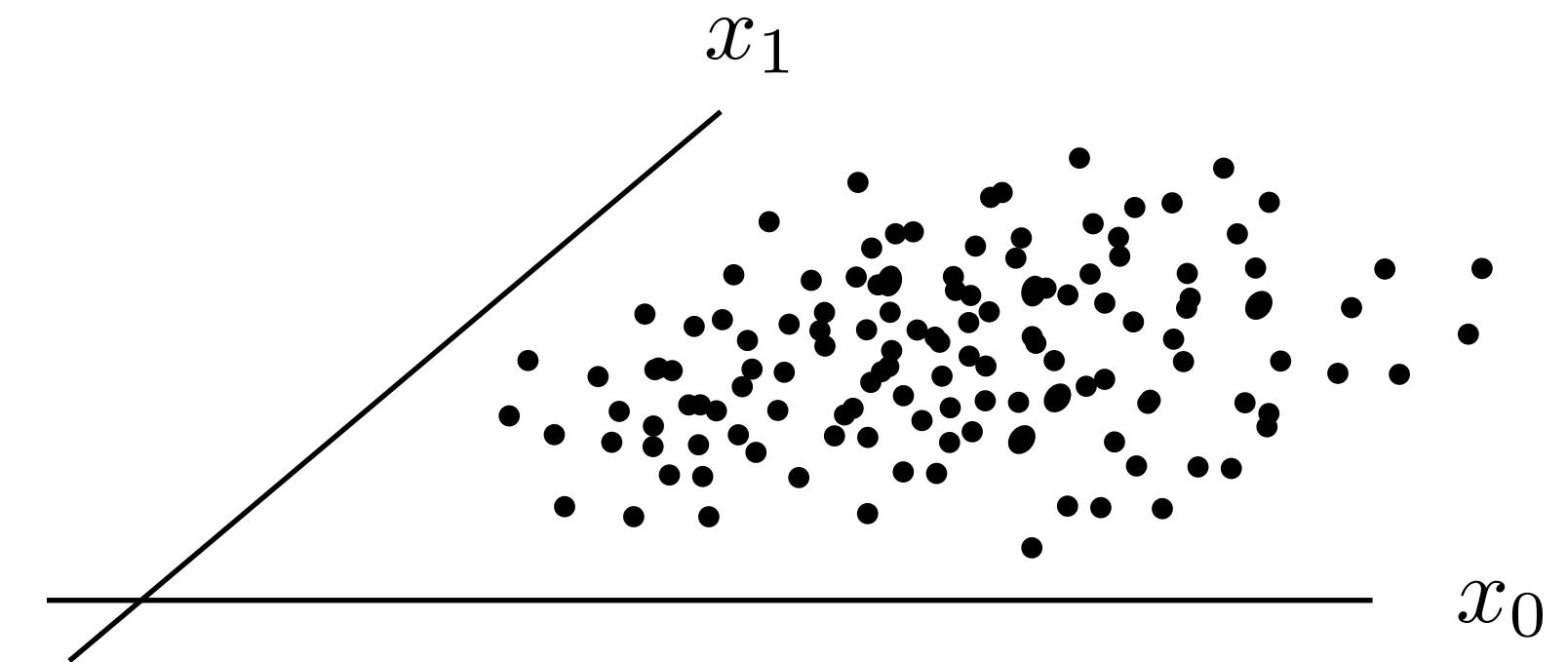


COST:

$$\min ||y - X\theta||^2$$

SOLN:

$$\theta = (X^T X)^{-1} X^T y$$



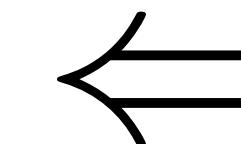
# Linear Regression

OUTPUTS  
(Dependent Variables)

$$y_t = \theta_t^T h_t(x_t)$$

INPUTS  
(Independent Variables)

$$\begin{bmatrix} y_{00} & \cdots & y_{0m} & \gamma_{00} & \cdots & \gamma_{0m'} \\ y_{10} & \cdots & y_{1m} & \gamma_{10} & \cdots & \gamma_{1m'} \\ y_{20} & \cdots & y_{2m} & \gamma_{20} & \cdots & \gamma_{2m'} \\ y_{30} & \cdots & y_{3m} & \gamma_{30} & \cdots & \gamma_{3m'} \\ y_{40} & \cdots & y_{4m} & \gamma_{40} & \cdots & \gamma_{4m'} \\ \vdots & & \vdots & \vdots & & \vdots \\ y_{T0} & \cdots & y_{Tm} & \gamma_{T0} & \cdots & \gamma_{Tm'} \end{bmatrix}$$



$$\begin{bmatrix} h_{00}(x_0) & \cdots & h_{0n}(x_0) \\ h_{10}(x_1) & \cdots & h_{1n}(x_1) \\ h_{20}(x_2) & \cdots & h_{2n}(x_2) \\ h_{30}(x_3) & \cdots & h_{3n}(x_3) \\ h_{40}(x_4) & \cdots & h_{4n}(x_4) \\ \vdots & & \vdots \\ h_{T0}(x_T) & \cdots & h_{Tn}(x_T) \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}$$

with BASIS  
FUNCTIONS

$$y = h(X)\theta$$

$y_0$

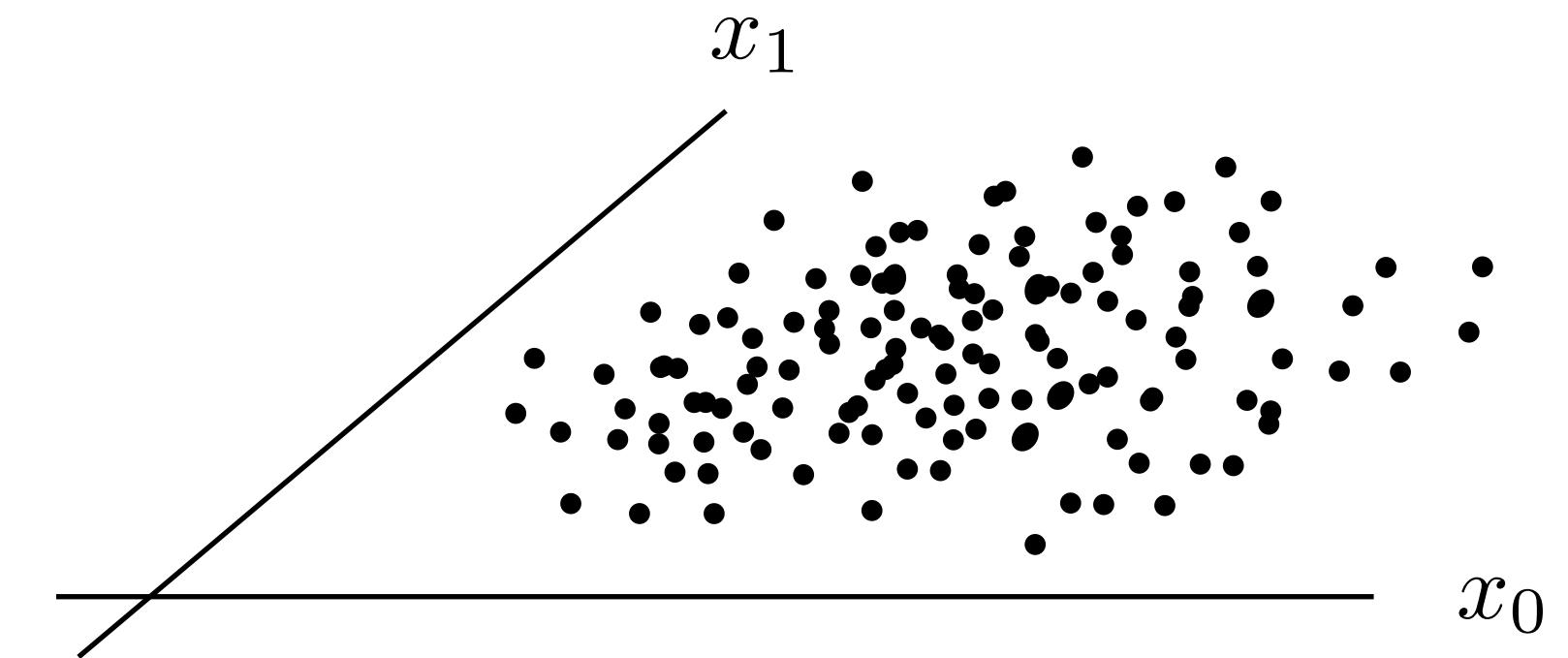


COST:

$$\min ||y - h(X)\theta||^2$$

SOLN:

$$\theta = (h(X)^T h(X))^{-1} h(X)^T y$$



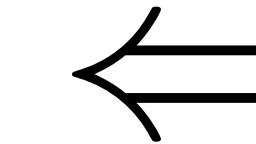
# Linear Regression

OUTPUTS  
(Dependent Variables)

$$y_t = \theta_t^T h_t(x_t)$$

INPUTS  
(Independent Variables)

$$\begin{bmatrix} y_{00} & \cdots & y_{0m} & \gamma_{00} & \cdots & \gamma_{0m'} \\ y_{10} & \cdots & y_{1m} & \gamma_{10} & \cdots & \gamma_{1m'} \\ y_{20} & \cdots & y_{2m} & \gamma_{20} & \cdots & \gamma_{2m'} \\ y_{30} & \cdots & y_{3m} & \gamma_{30} & \cdots & \gamma_{3m'} \\ y_{40} & \cdots & y_{4m} & \gamma_{40} & \cdots & \gamma_{4m'} \\ \vdots & & \vdots & \vdots & & \vdots \\ y_{T0} & \cdots & y_{Tm} & \gamma_{T0} & \cdots & \gamma_{Tm'} \end{bmatrix}$$



$$\begin{bmatrix} 1 & x_{00} & x_{00}^2 & x_{00}^3 \\ 1 & x_{10} & x_{10}^2 & x_{10}^3 \\ 1 & x_{20} & x_{20}^2 & x_{20}^3 \\ 1 & x_{30} & x_{30}^2 & x_{30}^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{T0} & x_{T0}^2 & x_{T0}^3 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}$$

with BASIS  
FUNCTIONS

$$y = h(X)\theta$$

$$h_t(x_t) = [1 \ x_{t0} \ x_{t0}^2 \ x_{t0}^3]$$

$y_0$



COST:

$$\min ||y - h(X)\theta||^2$$

SOLN:

$$\theta = (h(X)^T h(X))^{-1} h(X)^T y$$



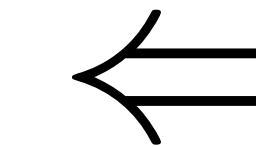
# Linear Regression

OUTPUTS  
(Dependent Variables)

$$y_t = \theta_t^T h_t(x_t)$$

INPUTS  
(Independent Variables)

$$\begin{bmatrix} y_{00} & \cdots & y_{0m} & \gamma_{00} & \cdots & \gamma_{0m'} \\ y_{10} & \cdots & y_{1m} & \gamma_{10} & \cdots & \gamma_{1m'} \\ y_{20} & \cdots & y_{2m} & \gamma_{20} & \cdots & \gamma_{2m'} \\ y_{30} & \cdots & y_{3m} & \gamma_{30} & \cdots & \gamma_{3m'} \\ y_{40} & \cdots & y_{4m} & \gamma_{40} & \cdots & \gamma_{4m'} \\ \vdots & & \vdots & \vdots & & \vdots \\ y_{T0} & \cdots & y_{Tm} & \gamma_{T0} & \cdots & \gamma_{Tm'} \end{bmatrix}$$

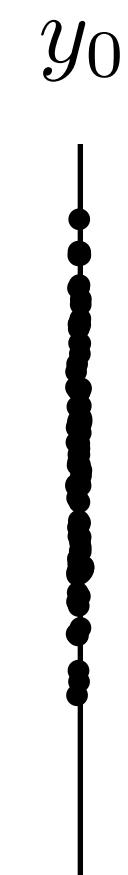


$$\begin{bmatrix} 1 & x_{00} & x_{00}^2 & x_{00}^3 \\ 1 & x_{10} & x_{10}^2 & x_{10}^3 \\ 1 & x_{20} & x_{20}^2 & x_{20}^3 \\ 1 & x_{30} & x_{30}^2 & x_{30}^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{T0} & x_{T0}^2 & x_{T0}^3 \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix}$$

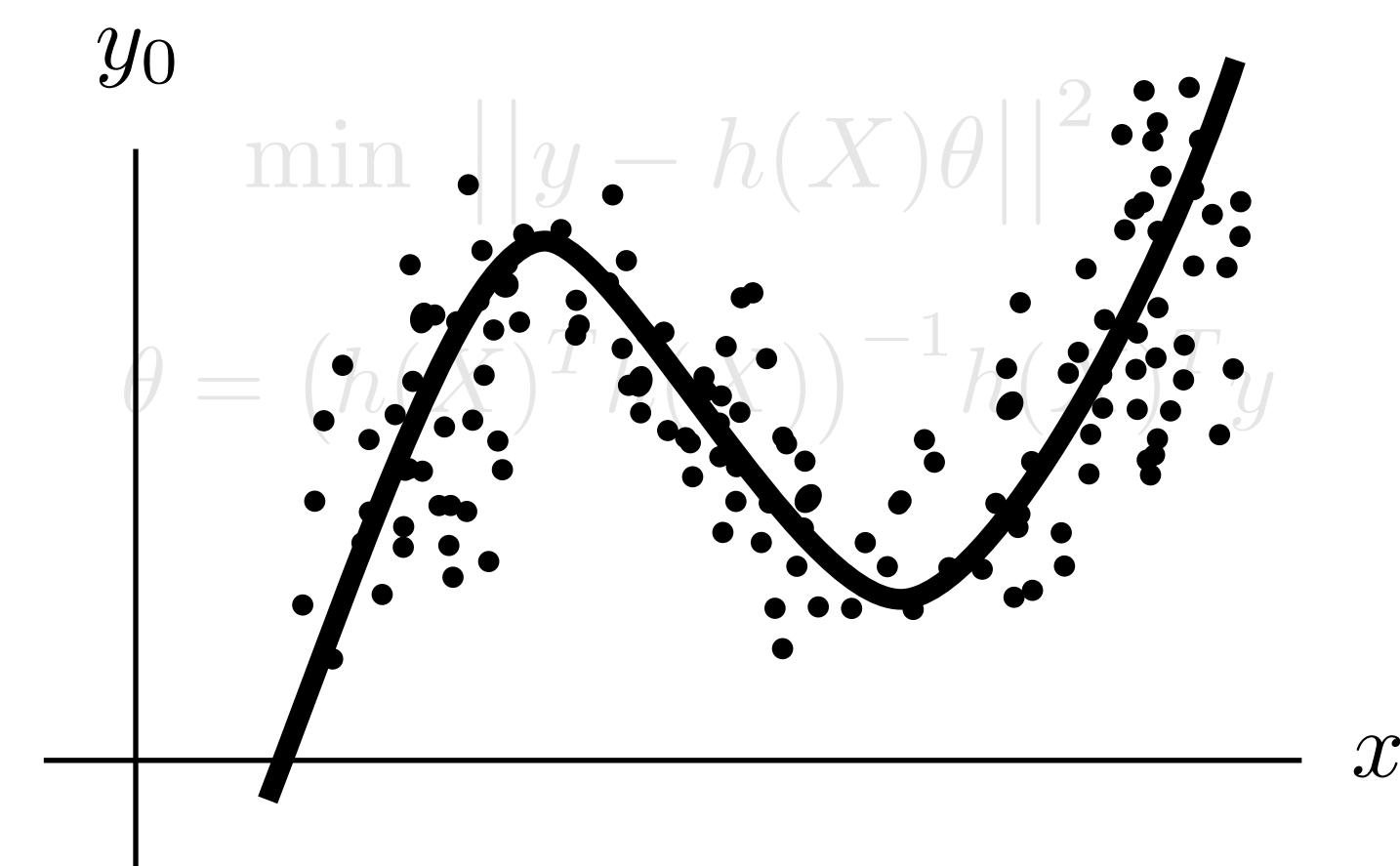
with BASIS  
FUNCTIONS

$$y = h(X)\theta$$

$$h_t(x_t) = [1 \ x_{t0} \ x_{t0}^2 \ x_{t0}^3]$$



COST:  
SOLN:



# Functions

OUTPUTS  
(Dependent Variables)

$$y_t = f(x_t)$$

INPUTS  
(Independent Variables)

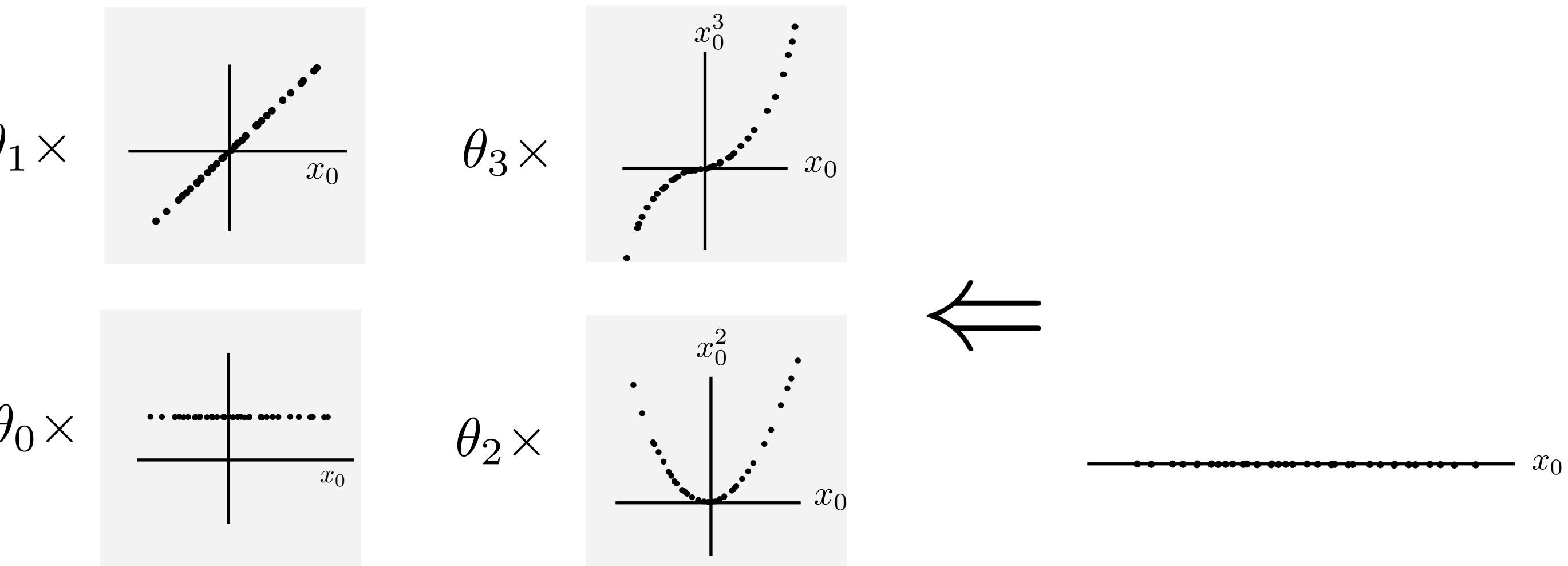
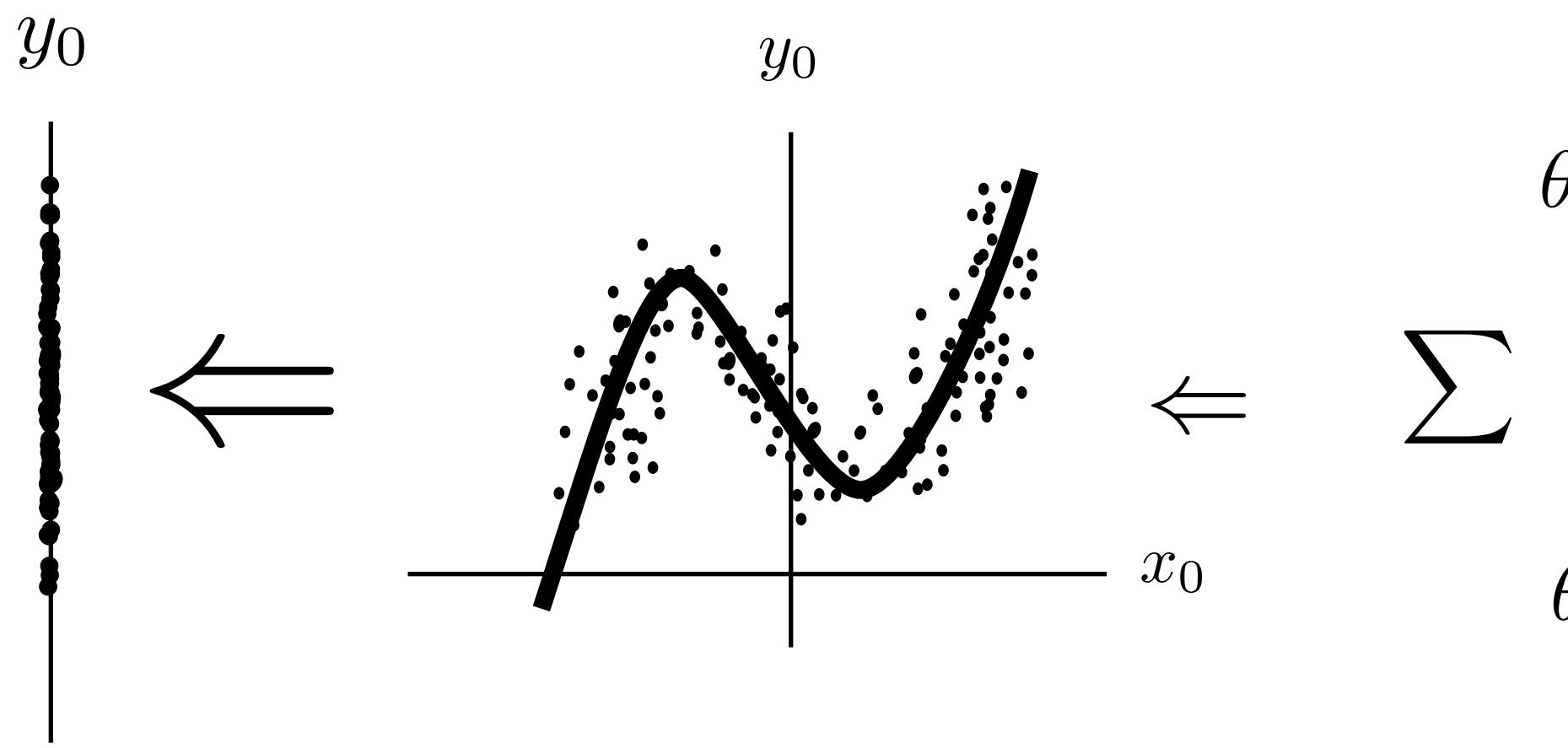
$$\begin{bmatrix} \gamma_{00} & \cdots & \gamma_{0m'} \\ \gamma_{10} & \cdots & \gamma_{1m'} \\ \gamma_{20} & \cdots & \gamma_{2m'} \\ \gamma_{30} & \cdots & \gamma_{3m'} \\ \gamma_{40} & \cdots & \gamma_{4m'} \\ \vdots & & \vdots \\ \gamma_{T0} & \cdots & \gamma_{Tm'} \end{bmatrix} \begin{bmatrix} y_{00} & \cdots & y_{0m} \\ y_{10} & \cdots & y_{1m} \\ y_{20} & \cdots & y_{2m} \\ y_{30} & \cdots & y_{3m} \\ y_{40} & \cdots & y_{4m} \\ \vdots & & \vdots \\ y_{T0} & \cdots & y_{Tm} \end{bmatrix} \leftarrow$$

$$\begin{bmatrix} h_{00}(x_0, \xi_0) & \cdots & h_{0n}(x_0, \xi_0) \\ h_{10}(x_1, \xi_1) & \cdots & h_{1n}(x_1, \xi_1) \\ h_{20}(x_2, \xi_2) & \cdots & h_{2n}(x_2, \xi_2) \\ h_{30}(x_3, \xi_3) & \cdots & h_{3n}(x_3, \xi_3) \\ h_{40}(x_4, \xi_4) & \cdots & h_{4n}(x_4, \xi_4) \\ \vdots & & \vdots \\ h_{T0}(x_T, \xi_T) & \cdots & h_{Tn}(x_T, \xi_T) \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{bmatrix} \leftarrow$$

$$f \begin{bmatrix} (x_{00} \cdots x_{0n}, \xi_{00} \cdots \xi_{0n'}) \\ x_{10} \cdots x_{1n}, \xi_{10} \cdots \xi_{1n'} \\ x_{20} \cdots x_{2n}, \xi_{20} \cdots \xi_{2n'} \\ x_{30} \cdots x_{3n}, \xi_{30} \cdots \xi_{3n'} \\ x_{40} \cdots x_{4n}, \xi_{40} \cdots \xi_{4n'} \\ \vdots & \vdots & \vdots & \vdots \\ x_{T0} \cdots x_{Tn}, \xi_{T0} \cdots \xi_{Tn'} \end{bmatrix}$$

$$h_t(x_t) = [1 \ x_{t0} \ x_{t0}^2 \ x_{t0}^3]$$

with BASIS FUNCTIONS



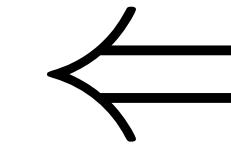
# Linear Regression

OUTPUTS  
(Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS  
(Independent Variables)

$$\begin{bmatrix} y_{00} & \cdots & y_{0m} & \gamma_{00} & \cdots & \gamma_{0m'} \\ y_{10} & \cdots & y_{1m} & \gamma_{10} & \cdots & \gamma_{1m'} \\ y_{20} & \cdots & y_{2m} & \gamma_{20} & \cdots & \gamma_{2m'} \\ y_{30} & \cdots & y_{3m} & \gamma_{30} & \cdots & \gamma_{3m'} \\ y_{40} & \cdots & y_{4m} & \gamma_{40} & \cdots & \gamma_{4m'} \\ \vdots & & \vdots & \vdots & & \vdots \\ y_{T0} & \cdots & y_{Tm} & \gamma_{T0} & \cdots & \gamma_{Tm'} \end{bmatrix}$$



$$\begin{bmatrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \\ \vdots \\ \xi_{T0} \end{bmatrix} \cdots \begin{bmatrix} \xi_{0n'} \\ \xi_{1n'} \\ \vdots \\ \xi_{2n'} \\ \vdots \\ \xi_{3n'} \\ \vdots \\ \xi_{4n'} \\ \vdots \\ \vdots \\ \xi_{Tn'} \end{bmatrix}$$

$$y = X\theta$$

$y_0$

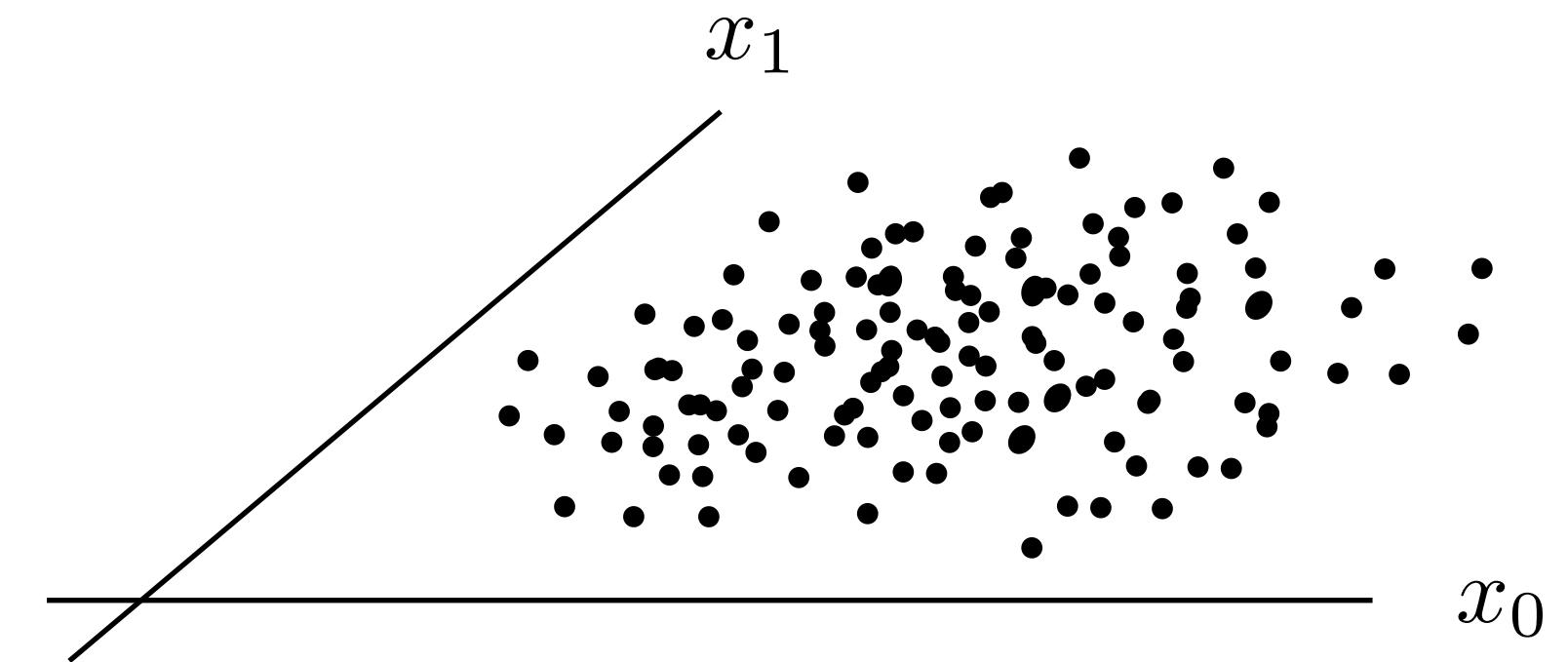


COST:

$$\min ||y - X\theta||^2$$

SOLN:

$$\theta = (X^T X)^{-1} X^T y$$



# Linear Regression

OUTPUTS  
(Dependent Variables)

$$y_t = \theta^T x_t + v_t$$

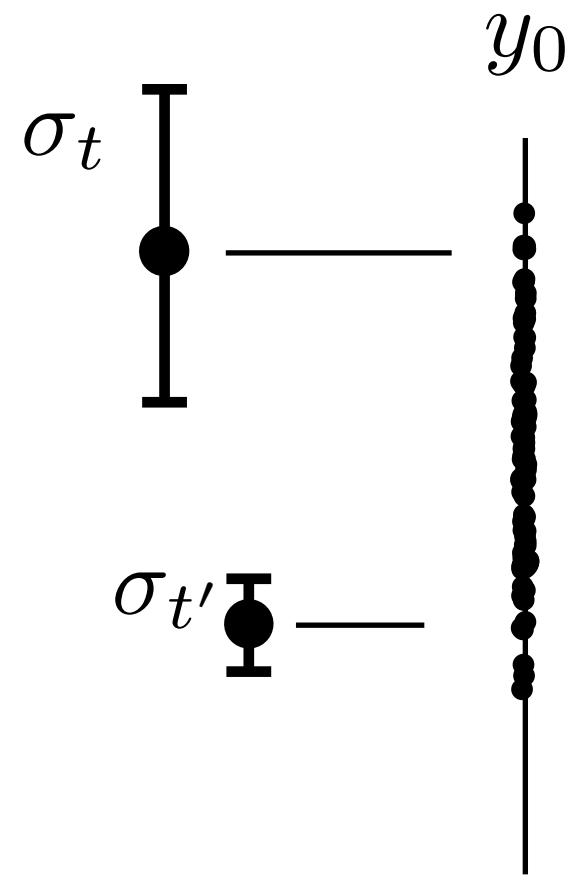
$$v_t \sim \mathcal{N}(0, \sigma_t^2)$$

INPUTS  
(Independent Variables)

$$\begin{bmatrix} y_{00} & \cdots & y_{0m} & \gamma_{00} & \cdots & \gamma_{0m'} \\ y_{10} & \cdots & y_{1m} & \gamma_{10} & \cdots & \gamma_{1m'} \\ y_{20} & \cdots & y_{2m} & \gamma_{20} & \cdots & \gamma_{2m'} \\ y_{30} & \cdots & y_{3m} & \gamma_{30} & \cdots & \gamma_{3m'} \\ y_{40} & \cdots & y_{4m} & \gamma_{40} & \cdots & \gamma_{4m'} \\ \vdots & & \vdots & \vdots & & \vdots \\ y_{T0} & \cdots & y_{Tm} & \gamma_{T0} & \cdots & \gamma_{Tm'} \end{bmatrix}$$

$$\begin{bmatrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \\ \vdots \\ \xi_{T0} \end{bmatrix} \cdots \begin{bmatrix} \xi_{0n'} \\ \xi_{1n'} \\ \xi_{2n'} \\ \vdots \\ \xi_{3n'} \\ \vdots \\ \xi_{4n'} \\ \vdots \\ \vdots \\ \xi_{Tn'} \end{bmatrix}$$

$$y = X\theta + v$$



**COST:**

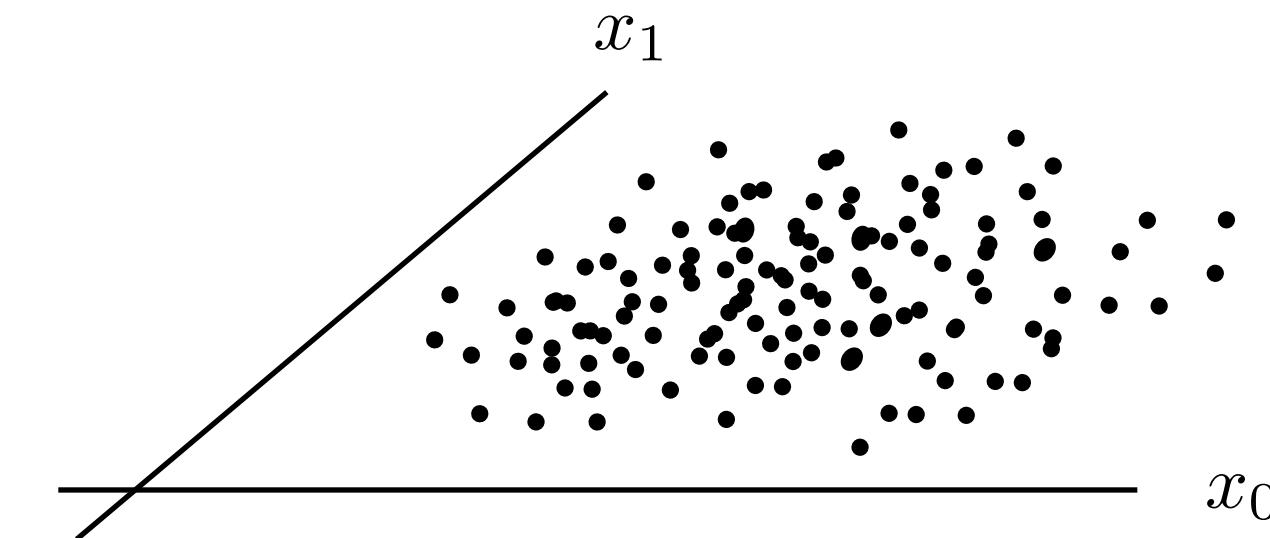
$$\min (y - X\theta)^T W^{-1} (y - X\theta)$$

**SOLN:**

$$\theta = (X^T W^{-1} X)^{-1} X W^{-1} y$$

**WEIGHTS:**  
Weighted  
Least Squares

$$W = \text{diag}(\sigma_0^2, \dots, \sigma_T^2)$$



# Linear Regression

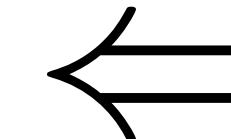
OUTPUTS  
(Dependent Variables)

$$y_t = \theta^T x_t + v_t$$

$$v_t \sim \mathcal{N}(0, \sigma_t^2)$$

INPUTS  
(Independent Variables)

$$\begin{bmatrix} y_{00} & \cdots & y_{0m} & \gamma_{00} & \cdots & \gamma_{0m'} \\ y_{10} & \cdots & y_{1m} & \gamma_{10} & \cdots & \gamma_{1m'} \\ y_{20} & \cdots & y_{2m} & \gamma_{20} & \cdots & \gamma_{2m'} \\ y_{30} & \cdots & y_{3m} & \gamma_{30} & \cdots & \gamma_{3m'} \\ y_{40} & \cdots & y_{4m} & \gamma_{40} & \cdots & \gamma_{4m'} \\ \vdots & & \vdots & \vdots & & \vdots \\ y_{T0} & \cdots & y_{Tm} & \gamma_{T0} & \cdots & \gamma_{Tm'} \end{bmatrix}$$

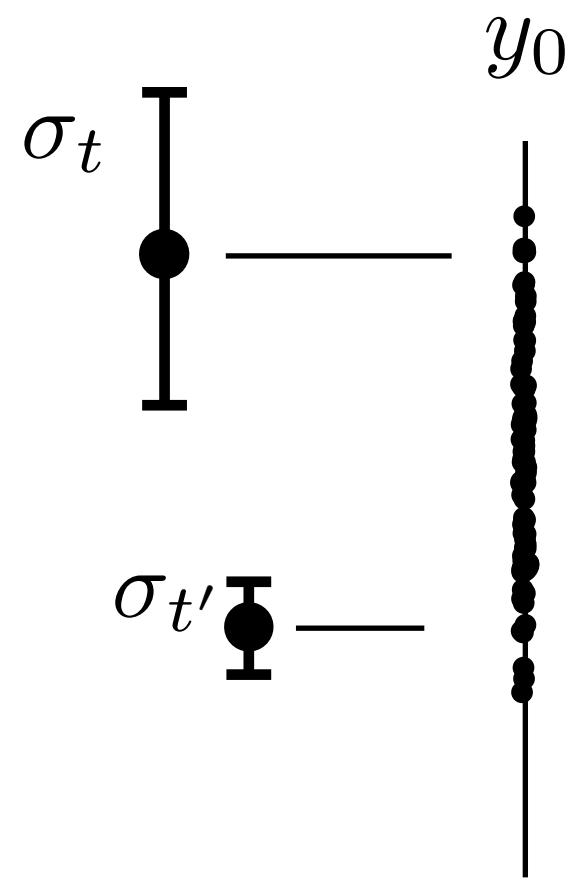


$$\begin{bmatrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \\ \vdots \\ \xi_{T0} \\ \cdots \\ \xi_{Tn'} \end{bmatrix}$$

Prior on

$$y = X\theta + v$$

$$\theta \sim \mathcal{N}(\mu, Q)$$



**COST:**

$$\min (y - X\theta)^T W^{-1} (y - X\theta) + (\theta - \mu)^T Q^{-1} (\theta - \mu)$$

**SOLN:**

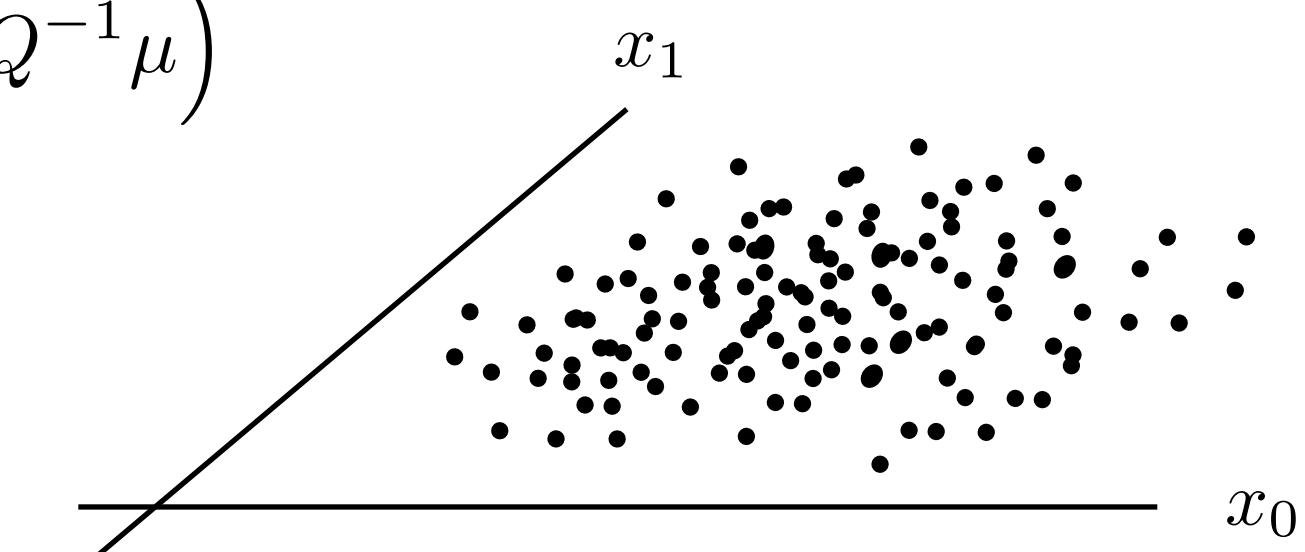
$$\theta = (X^T W^{-1} X + Q^{-1})^{-1} (X^T W^{-1} y + Q^{-1} \mu)$$

**WEIGHTS:**

$$W = \text{diag}(\sigma_0^2, \dots, \sigma_T^2)$$

Bayesian  
Ridge Regression

$$Q = \lambda^{-1} I$$



# Linear Regression

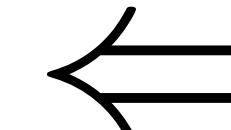
OUTPUTS  
(Dependent Variables)

$$y_t = \theta^T x_t + v_t$$

$$v_t \sim \mathcal{N}(0, \sigma_t^2)$$

INPUTS  
(Independent Variables)

$$\begin{bmatrix} y_{00} & \cdots & y_{0m} & \gamma_{00} & \cdots & \gamma_{0m'} \\ y_{10} & \cdots & y_{1m} & \gamma_{10} & \cdots & \gamma_{1m'} \\ y_{20} & \cdots & y_{2m} & \gamma_{20} & \cdots & \gamma_{2m'} \\ y_{30} & \cdots & y_{3m} & \gamma_{30} & \cdots & \gamma_{3m'} \\ y_{40} & \cdots & y_{4m} & \gamma_{40} & \cdots & \gamma_{4m'} \\ \vdots & & \vdots & \vdots & & \vdots \\ y_{T0} & \cdots & y_{Tm} & \gamma_{T0} & \cdots & \gamma_{Tm'} \end{bmatrix}$$

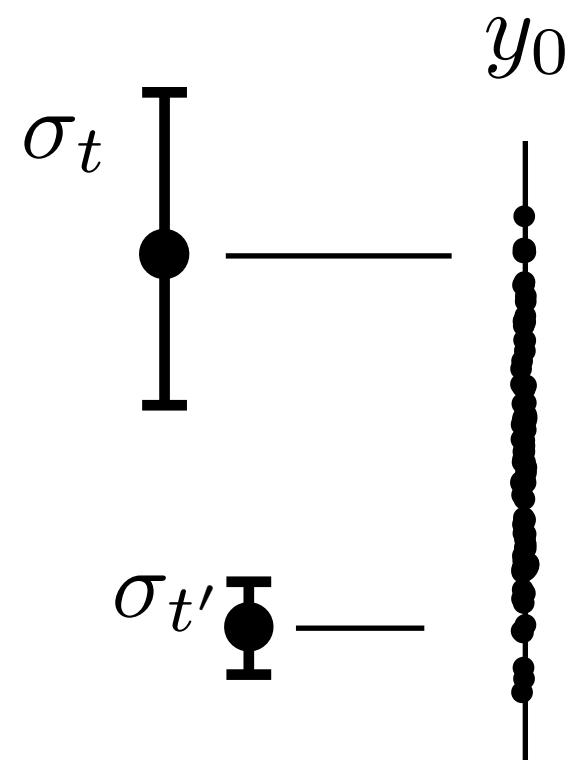


$$\begin{bmatrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \\ \vdots \\ \xi_{T0} \\ \cdots \\ \xi_{Tn'} \end{bmatrix}$$

Prior on

$$y = X\theta + v$$

$$\theta \sim \mathcal{N}(\mu, Q)$$



**COST:**

$$\min (y - X\theta)^T W^{-1} (y - X\theta) + (\theta - \mu)^T Q^{-1} (\theta - \mu)$$

**SOLN:**

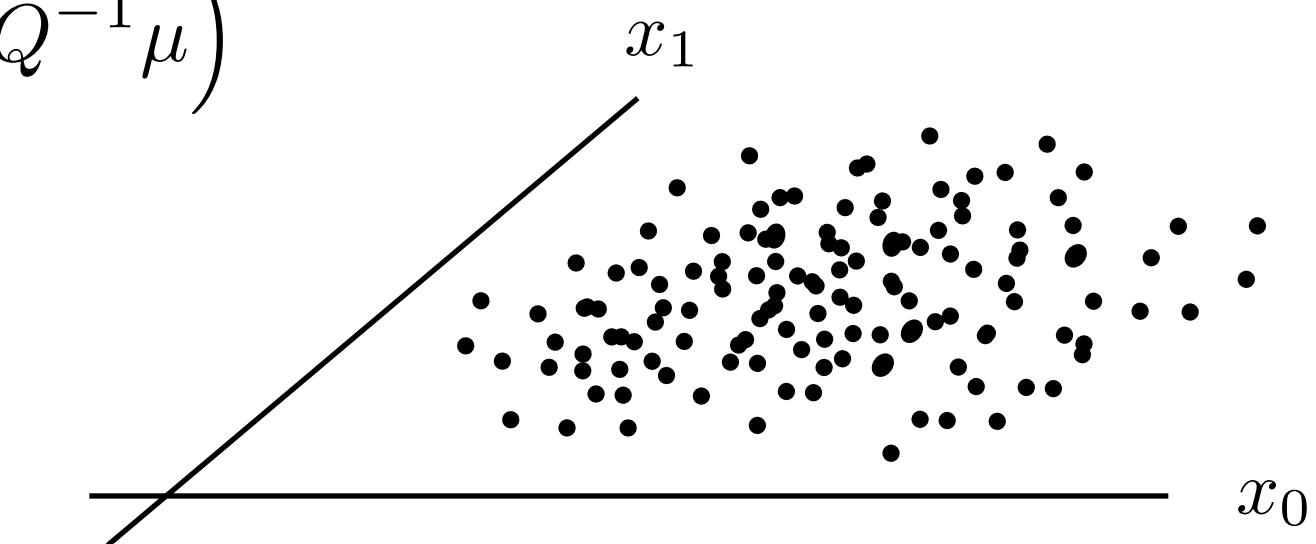
$$\theta = (X^T W^{-1} X + Q^{-1})^{-1} (X^T W^{-1} y + Q^{-1} \mu)$$

**WEIGHTS:**

$$W = \text{diag}(\sigma_0^2, \dots, \sigma_T^2)$$

Automatic Relevance  
Determination (ARD)

$$Q = \text{diag}(\lambda_0, \dots, \lambda_n)^{-1}$$



# Linear Regression

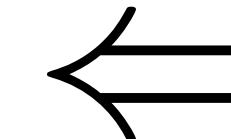
OUTPUTS  
(Dependent Variables)

$$y_t = \theta^T x_t + v_t$$

$$v_t \sim \mathcal{N}(0, \sigma_t^2)$$

INPUTS  
(Independent Variables)

$$\begin{bmatrix} y_{00} & \cdots & y_{0m} & \gamma_{00} & \cdots & \gamma_{0m'} \\ y_{10} & \cdots & y_{1m} & \gamma_{10} & \cdots & \gamma_{1m'} \\ y_{20} & \cdots & y_{2m} & \gamma_{20} & \cdots & \gamma_{2m'} \\ y_{30} & \cdots & y_{3m} & \gamma_{30} & \cdots & \gamma_{3m'} \\ y_{40} & \cdots & y_{4m} & \gamma_{40} & \cdots & \gamma_{4m'} \\ \vdots & & \vdots & \vdots & & \vdots \\ y_{T0} & \cdots & y_{Tm} & \gamma_{T0} & \cdots & \gamma_{Tm'} \end{bmatrix}$$

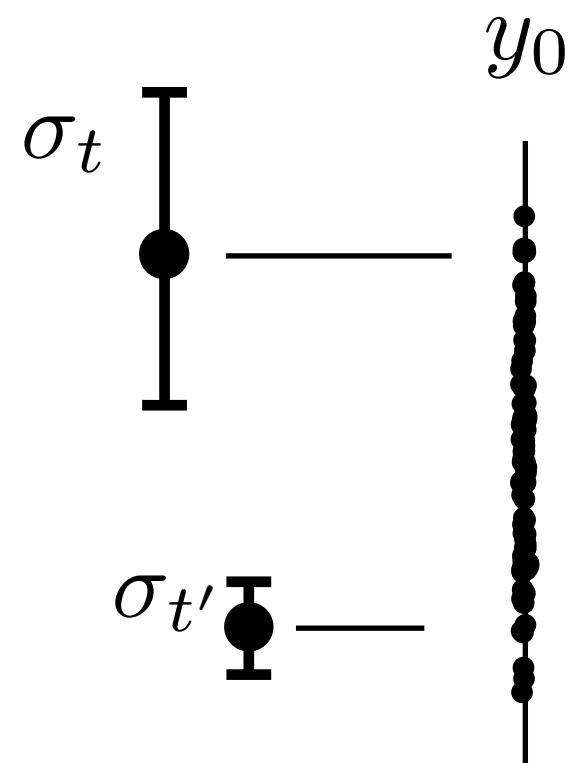


$$\begin{bmatrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \\ \vdots \\ \xi_{T0} \\ \cdots \\ \xi_{Tn'} \end{bmatrix}$$

Prior on

$$y = X\theta + v$$

$$\theta \sim \mathcal{N}(\mu, Q)$$



**COST:**

$$\min (y - X\theta)^T W^{-1} (y - X\theta) + (\theta - \mu)^T Q^{-1} (\theta - \mu)$$

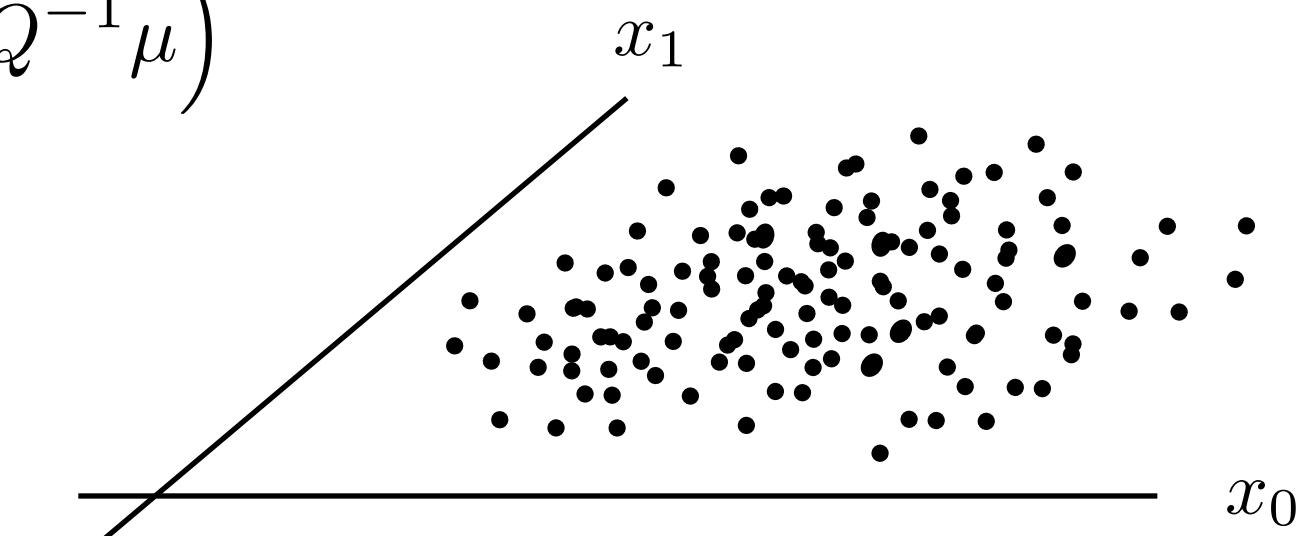
**SOLN:**

$$\theta = (X^T W^{-1} X + Q^{-1})^{-1} (X^T W^{-1} y + Q^{-1} \mu)$$

**WEIGHTS:**  
General  
Bayesian  
Regression

$$W = \text{diag}(\sigma_0^2, \dots, \sigma_T^2)$$

$$Q \succ 0$$



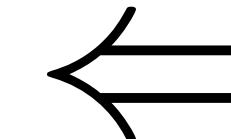
# Linear Regression

OUTPUTS  
(Dependent Variables)

$$y_t = \theta^T x_t$$

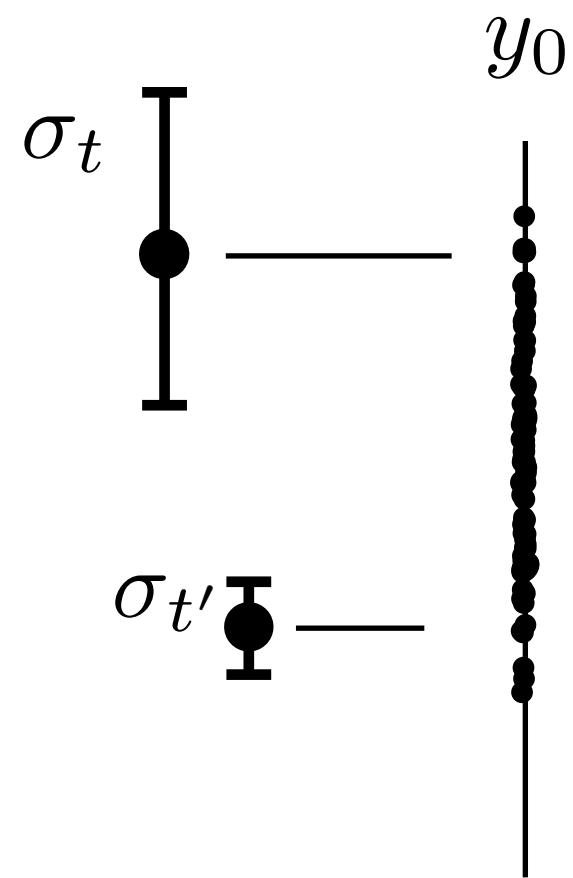
INPUTS  
(Independent Variables)

$$\begin{bmatrix} y_{00} & \cdots & y_{0m} & \gamma_{00} & \cdots & \gamma_{0m'} \\ y_{10} & \cdots & y_{1m} & \gamma_{10} & \cdots & \gamma_{1m'} \\ y_{20} & \cdots & y_{2m} & \gamma_{20} & \cdots & \gamma_{2m'} \\ y_{30} & \cdots & y_{3m} & \gamma_{30} & \cdots & \gamma_{3m'} \\ y_{40} & \cdots & y_{4m} & \gamma_{40} & \cdots & \gamma_{4m'} \\ \vdots & & \vdots & \vdots & & \vdots \\ y_{T0} & \cdots & y_{Tm} & \gamma_{T0} & \cdots & \gamma_{Tm'} \end{bmatrix}$$



$$\begin{bmatrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{bmatrix} \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \\ \vdots \\ \xi_{T0} \end{bmatrix} \cdots \begin{bmatrix} \xi_{0n'} \\ \xi_{1n'} \\ \xi_{2n'} \\ \vdots \\ \xi_{3n'} \\ \vdots \\ \xi_{4n'} \\ \vdots \\ \vdots \\ \xi_{Tn'} \end{bmatrix}$$

$$y = X\theta$$

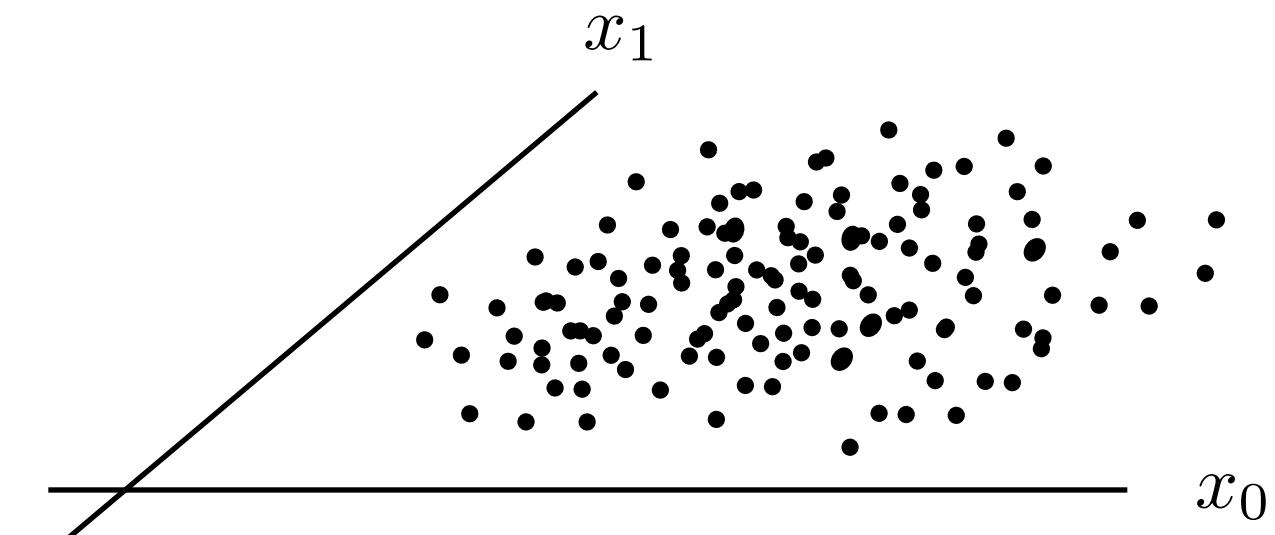


**COST:**  
**REGULARIZE:**  
Ridge  
Regression

$$\min ||y - X\theta||^2 + \rho(\theta)$$

$$\rho(\theta) \propto ||\theta||_2^2$$

“small  $\theta$ ”



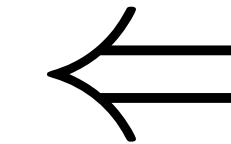
# Linear Regression

OUTPUTS  
(Dependent Variables)

$$y_t = \theta^T x_t$$

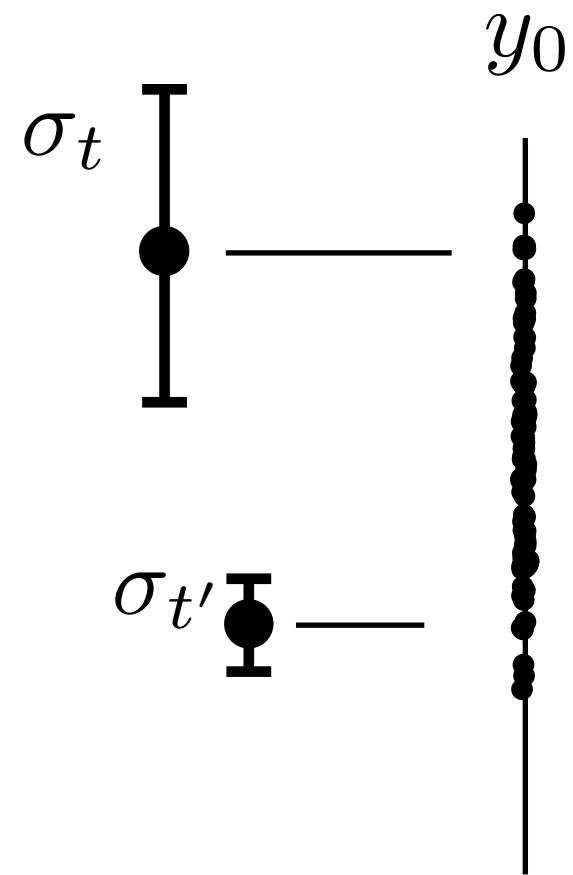
INPUTS  
(Independent Variables)

$$\begin{bmatrix} y_{00} & \cdots & y_{0m} & \gamma_{00} & \cdots & \gamma_{0m'} \\ y_{10} & \cdots & y_{1m} & \gamma_{10} & \cdots & \gamma_{1m'} \\ y_{20} & \cdots & y_{2m} & \gamma_{20} & \cdots & \gamma_{2m'} \\ y_{30} & \cdots & y_{3m} & \gamma_{30} & \cdots & \gamma_{3m'} \\ y_{40} & \cdots & y_{4m} & \gamma_{40} & \cdots & \gamma_{4m'} \\ \vdots & & \vdots & \vdots & & \vdots \\ y_{T0} & \cdots & y_{Tm} & \gamma_{T0} & \cdots & \gamma_{Tm'} \end{bmatrix}$$



$$\begin{bmatrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{bmatrix} \quad \begin{bmatrix} 0 \\ \theta_1 \\ \vdots \\ 0 \\ \theta_n \\ \vdots \\ \xi_{T0} \\ \cdots \\ \xi_{Tn'} \end{bmatrix}$$

$$y = X\theta$$



**COST:**

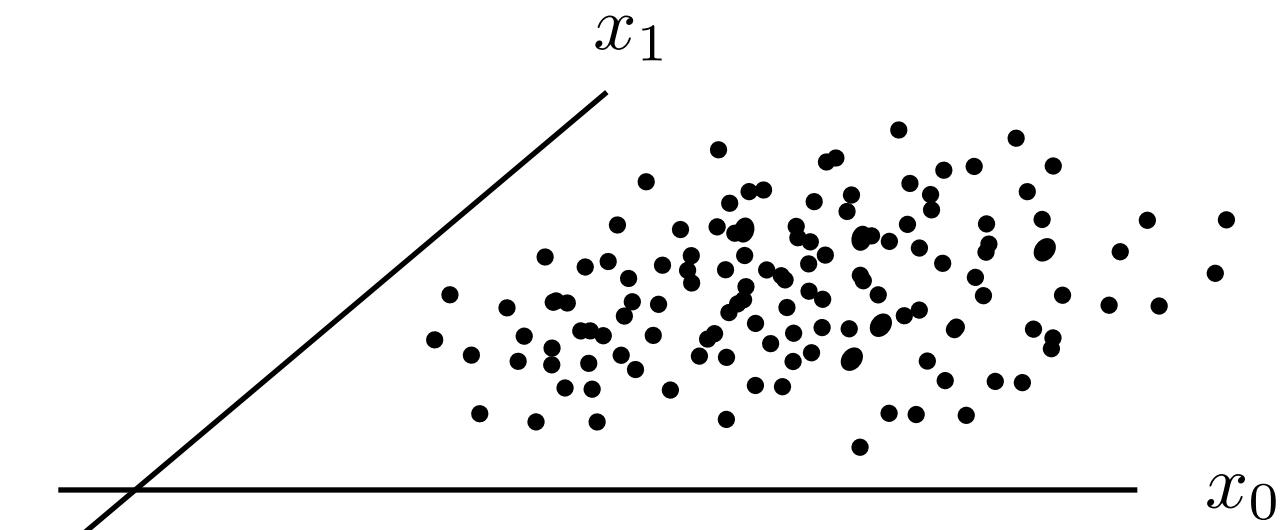
**REGULARIZE:**

- L1-LASSO
- Orthogonal Matching Pursuit (OMP)

$$\min \|y - X\theta\|^2 + \rho(\theta)$$

$$\rho(\theta) \propto \|\theta\|_1 \quad \text{"Sparsity"}$$

"Sparsity limit  
- greedy"



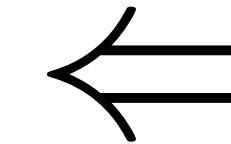
# Linear Regression

OUTPUTS  
(Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS  
(Independent Variables)

$$\begin{bmatrix} y_{00} & \cdots & y_{0m} \\ y_{10} & \cdots & y_{1m} \\ y_{20} & \cdots & y_{2m} \\ y_{30} & \cdots & y_{3m} \\ y_{40} & \cdots & y_{4m} \\ \vdots & & \vdots \\ y_{T0} & \cdots & y_{Tm} \end{bmatrix} \quad \begin{bmatrix} \gamma_{00} & \cdots & \gamma_{0m'} \\ \gamma_{10} & \cdots & \gamma_{1m'} \\ \gamma_{20} & \cdots & \gamma_{2m'} \\ \gamma_{30} & \cdots & \gamma_{3m'} \\ \gamma_{40} & \cdots & \gamma_{4m'} \\ \vdots & & \vdots \\ \gamma_{T0} & \cdots & \gamma_{Tm'} \end{bmatrix}$$



$$\begin{bmatrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{bmatrix} \quad \begin{bmatrix} \theta_{00} & \cdots & \theta_{0n} \\ \theta_{10} & \cdots & \theta_{1n} \\ \theta_{20} & \cdots & \theta_{2n} \\ \vdots & \ddots & \vdots \\ \theta_{30} & \cdots & \theta_{3n} \\ 0 & \cdots & 0 \\ \theta_{40} & \cdots & \theta_{4n} \\ \vdots & & \vdots \\ \theta_{n0} & \cdots & \theta_{nm} \end{bmatrix} \quad \begin{bmatrix} \xi_{T0} & \cdots & \xi_{Tn'} \end{bmatrix}$$

$$y = X\theta$$

COST:

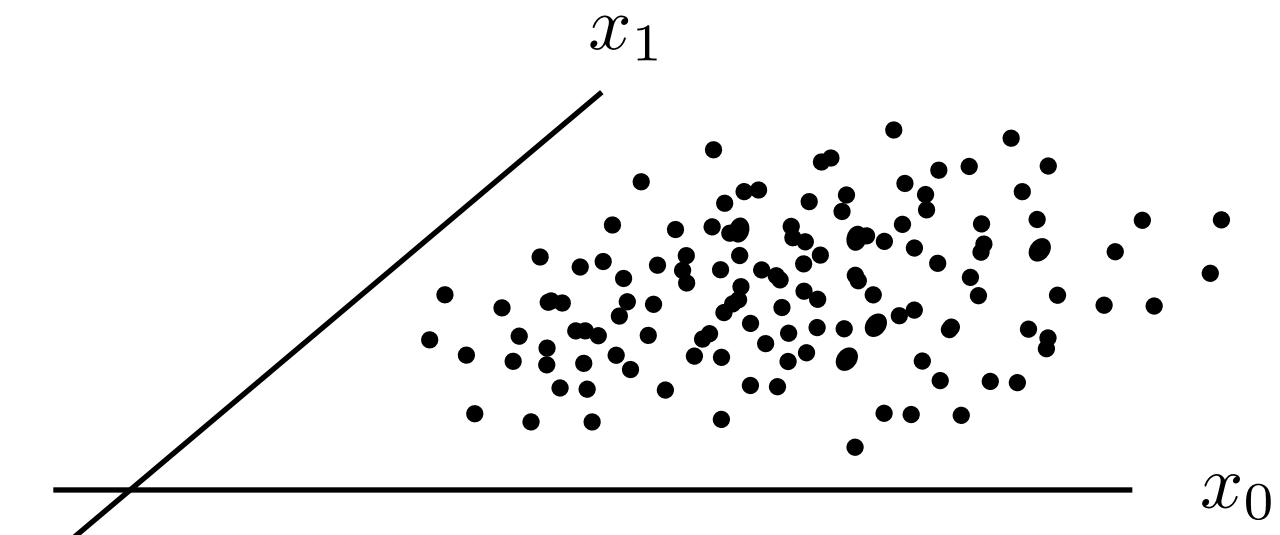
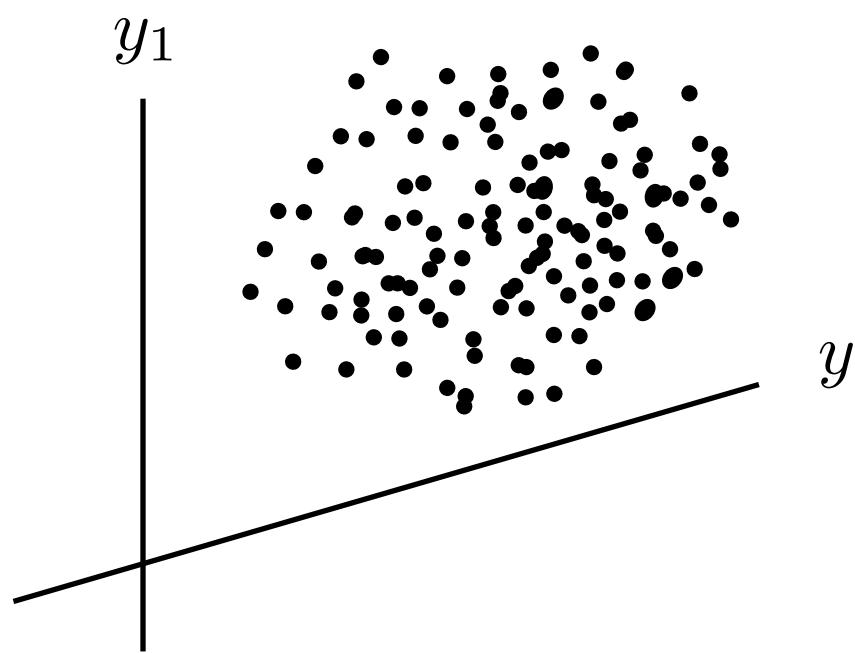
$$\min ||y - X\theta||_{\text{FRO}}^2 + \rho(\theta)$$

REGULARIZE:

Multi-task  
LASSO

$$\rho(\theta) \propto \sum_i ||\theta_{i:}||$$

“Sparsity across tasks”



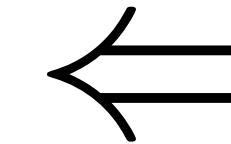
# Linear Regression

OUTPUTS  
(Dependent Variables)

$$y_t = \theta^T x_t$$

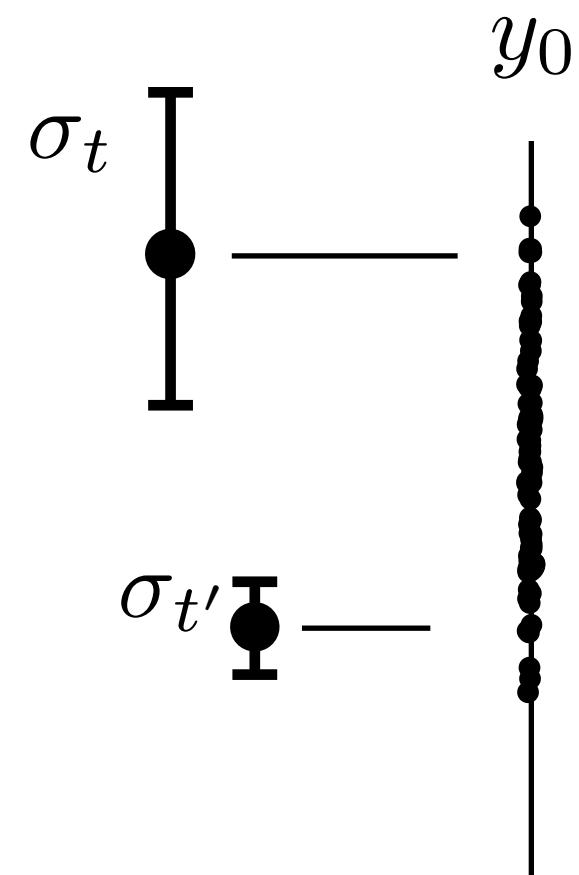
INPUTS  
(Independent Variables)

$$\begin{bmatrix} y_{00} & \cdots & y_{0m} & \gamma_{00} & \cdots & \gamma_{0m'} \\ y_{10} & \cdots & y_{1m} & \gamma_{10} & \cdots & \gamma_{1m'} \\ y_{20} & \cdots & y_{2m} & \gamma_{20} & \cdots & \gamma_{2m'} \\ y_{30} & \cdots & y_{3m} & \gamma_{30} & \cdots & \gamma_{3m'} \\ y_{40} & \cdots & y_{4m} & \gamma_{40} & \cdots & \gamma_{4m'} \\ \vdots & & \vdots & \vdots & & \vdots \\ y_{T0} & \cdots & y_{Tm} & \gamma_{T0} & \cdots & \gamma_{Tm'} \end{bmatrix}$$



$$\begin{bmatrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{bmatrix} \quad \begin{bmatrix} 0 \\ \theta_1 \\ \vdots \\ 0 \\ \theta_n \\ \vdots \\ \xi_{T0} \\ \cdots \\ \xi_{Tn'} \end{bmatrix}$$

$$y = X\theta$$



COST:

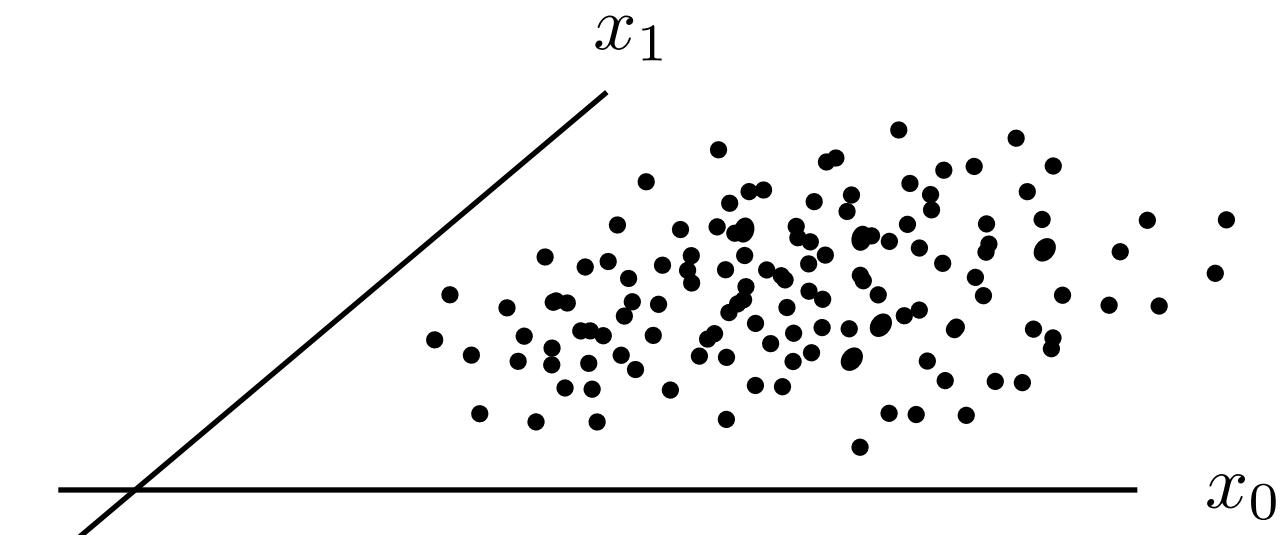
$$\min ||y - X\theta||^2 + \rho(\theta)$$

REGULARIZE:

Elastic-Net

$$\rho(\theta) \propto \beta ||\theta||_1 + \frac{1}{2}(1 - \beta) ||\theta||_2^2$$

"Mix of L1 and L2"



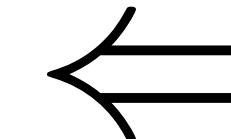
# Linear Regression

OUTPUTS  
(Dependent Variables)

$$y_t = \theta^T x_t$$

INPUTS  
(Independent Variables)

$$\begin{bmatrix} y_{00} & \cdots & y_{0m} \\ y_{10} & \cdots & y_{1m} \\ y_{20} & \cdots & y_{2m} \\ y_{30} & \cdots & y_{3m} \\ y_{40} & \cdots & y_{4m} \\ \vdots & & \vdots \\ y_{T0} & \cdots & y_{Tm} \end{bmatrix} \quad \begin{bmatrix} \gamma_{00} & \cdots & \gamma_{0m'} \\ \gamma_{10} & \cdots & \gamma_{1m'} \\ \gamma_{20} & \cdots & \gamma_{2m'} \\ \gamma_{30} & \cdots & \gamma_{3m'} \\ \gamma_{40} & \cdots & \gamma_{4m'} \\ \vdots & & \vdots \\ \gamma_{T0} & \cdots & \gamma_{Tm'} \end{bmatrix}$$



$$\begin{bmatrix} x_{00} & \cdots & x_{0n} \\ x_{10} & \cdots & x_{1n} \\ x_{20} & \cdots & x_{2n} \\ x_{30} & \cdots & x_{3n} \\ x_{40} & \cdots & x_{4n} \\ \vdots & & \vdots \\ x_{T0} & \cdots & x_{Tn} \end{bmatrix} \quad \begin{bmatrix} \theta_{00} & \cdots & \theta_{0n} \\ \theta_{10} & \cdots & \theta_{1n} \\ \theta_{20} & \cdots & \theta_{2n} \\ \vdots & \ddots & \vdots \\ \theta_{30} & \cdots & \theta_{3n} \\ 0 & \cdots & 0 \\ \theta_{40} & \cdots & \theta_{4n} \\ \vdots & & \vdots \\ \theta_{n0} & \cdots & \theta_{nm} \end{bmatrix} \quad \begin{bmatrix} \xi_{T0} & \cdots & \xi_{Tn'} \end{bmatrix}$$

$$y = X\theta$$

COST:

$$\min ||y - X\theta||_{\text{FRO}}^2 + \rho(\theta)$$

REGULARIZE:

Multi-Task  
Elastic-Net

$$\rho(\theta) \propto \beta \sum_i \|\theta_{i:}\|_2 + \frac{1}{2}(1-\beta)\|\theta\|_{\text{FRO}}^2$$

“Mixed L1 & L2 across tasks”

