MODEL FITTING II

Prof. Alexander Huth
9/21/2017

RECAP

* Linear regression, L2 loss (minimizing squared error)

RECAP

* Regularization
  * Prior
  * Penalty
  * Geometry / ad hoc (early stopping)

RECAP

* Families of regularized models
  * L2 / ridge / grad. descent w/ early stopping
  * L1 / LASSO / coord. descent w/ early stopping
  * L0 / feature selection

1D EXAMPLE

subject

stimulus

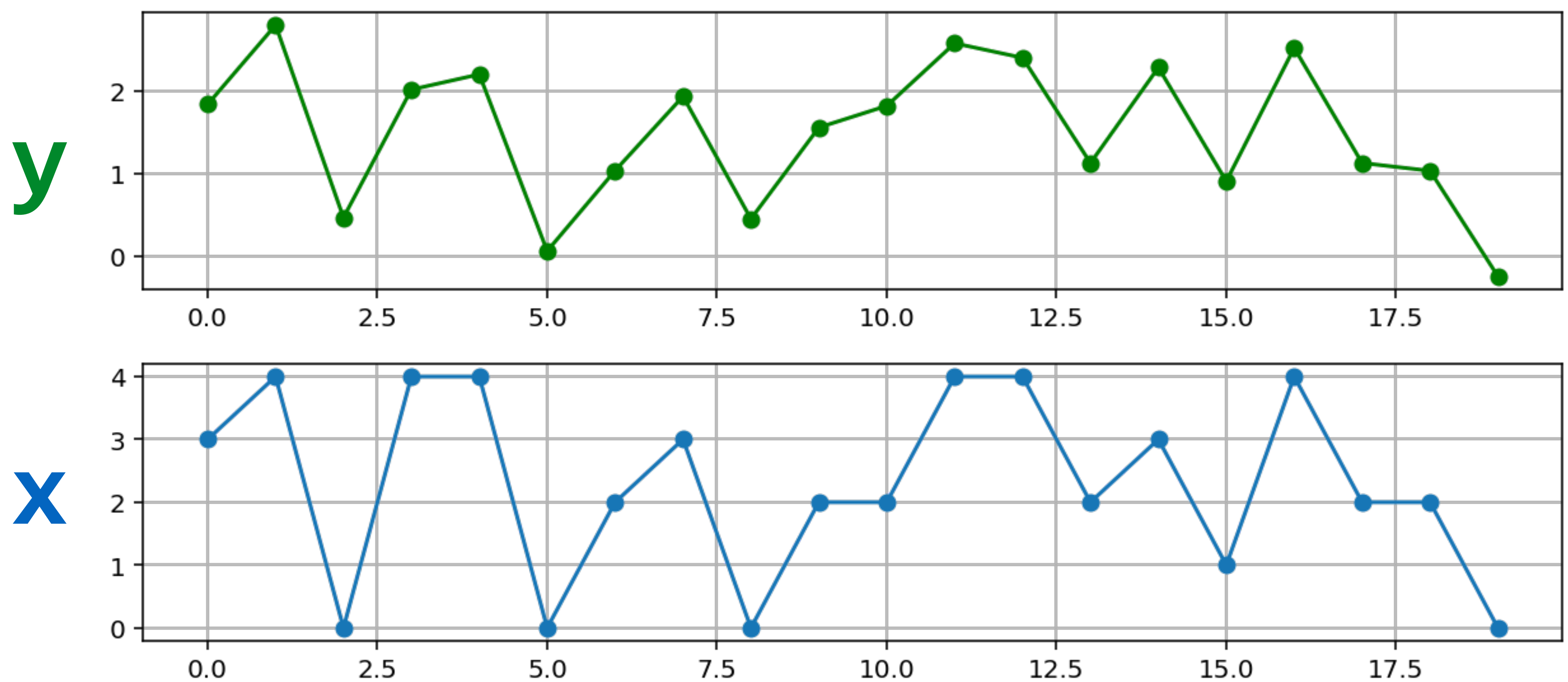
* **y** = output of a neuron that you are measuring

* **x** = how many times per second the screen flashes

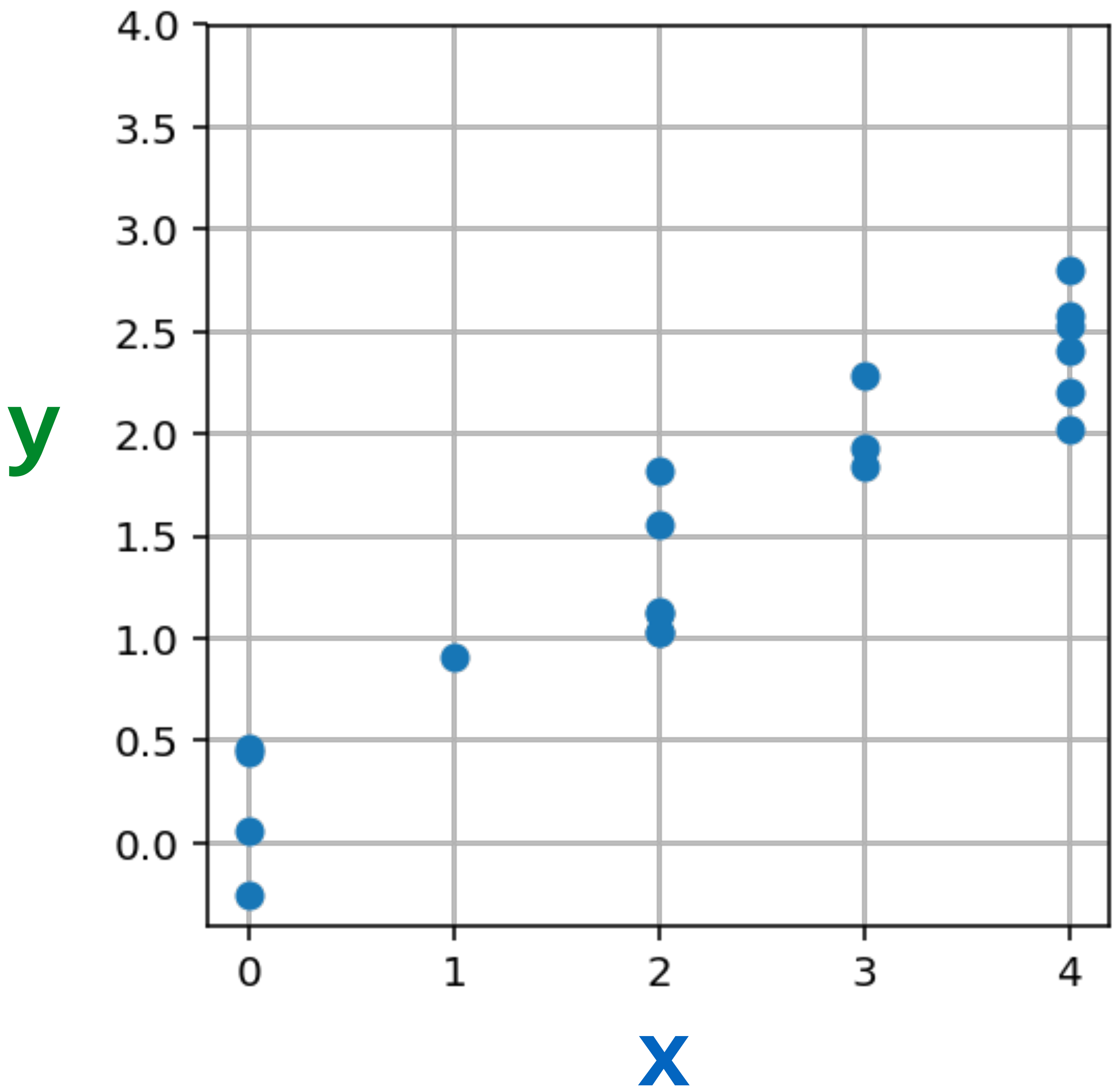
1D EXAMPLE

subject

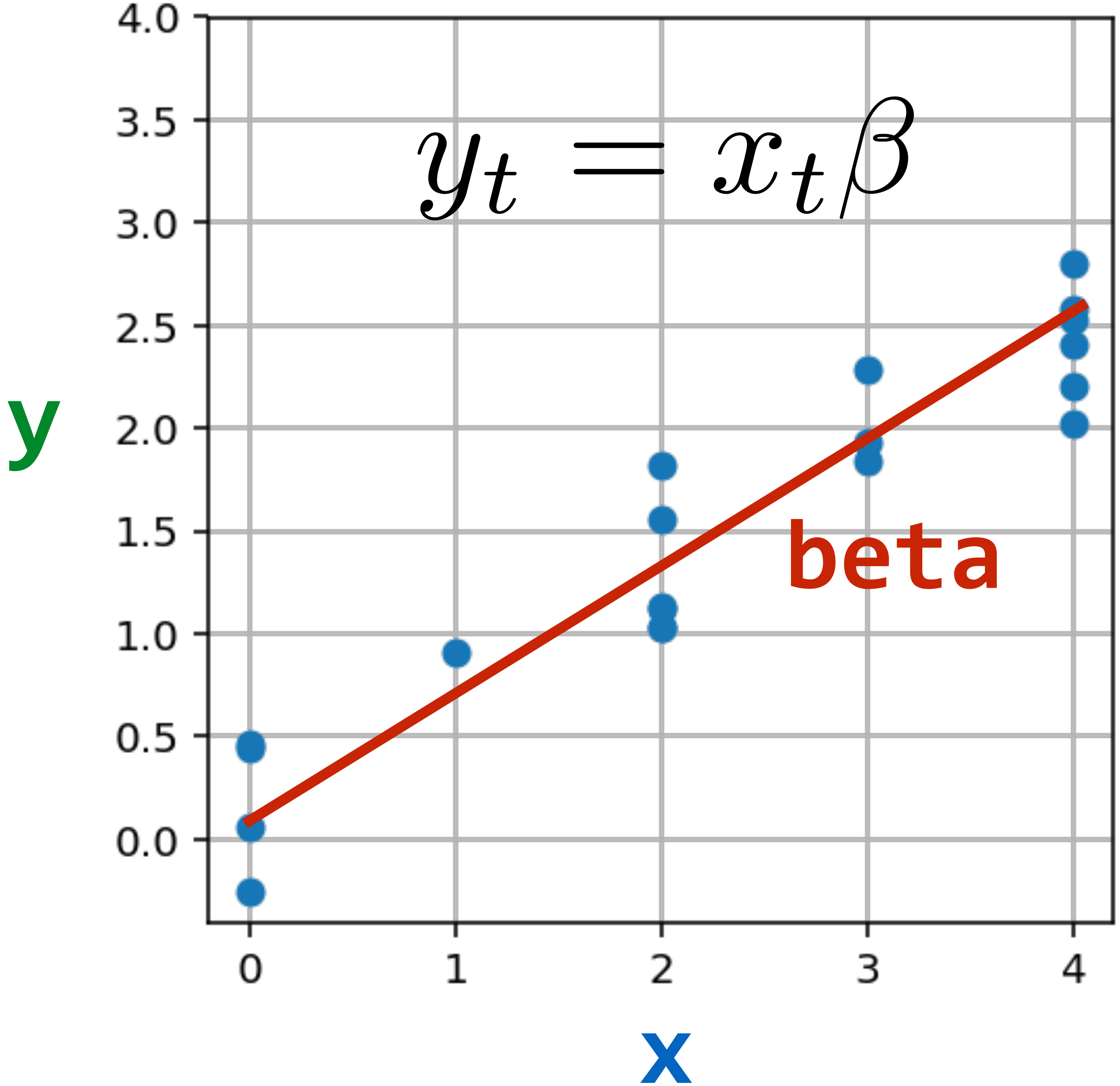
stimulus



1D EXAMPLE



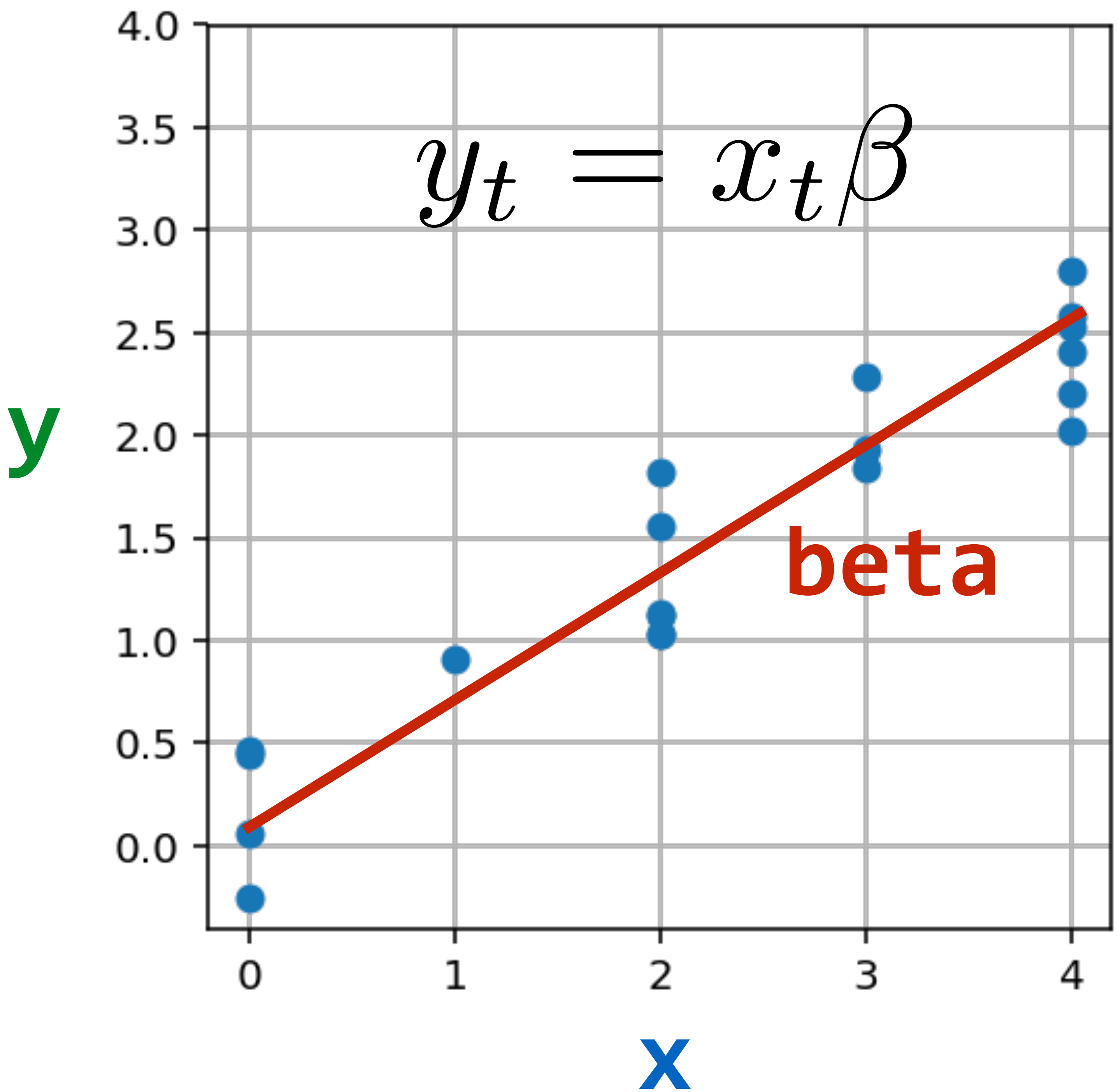
1D EXAMPLE

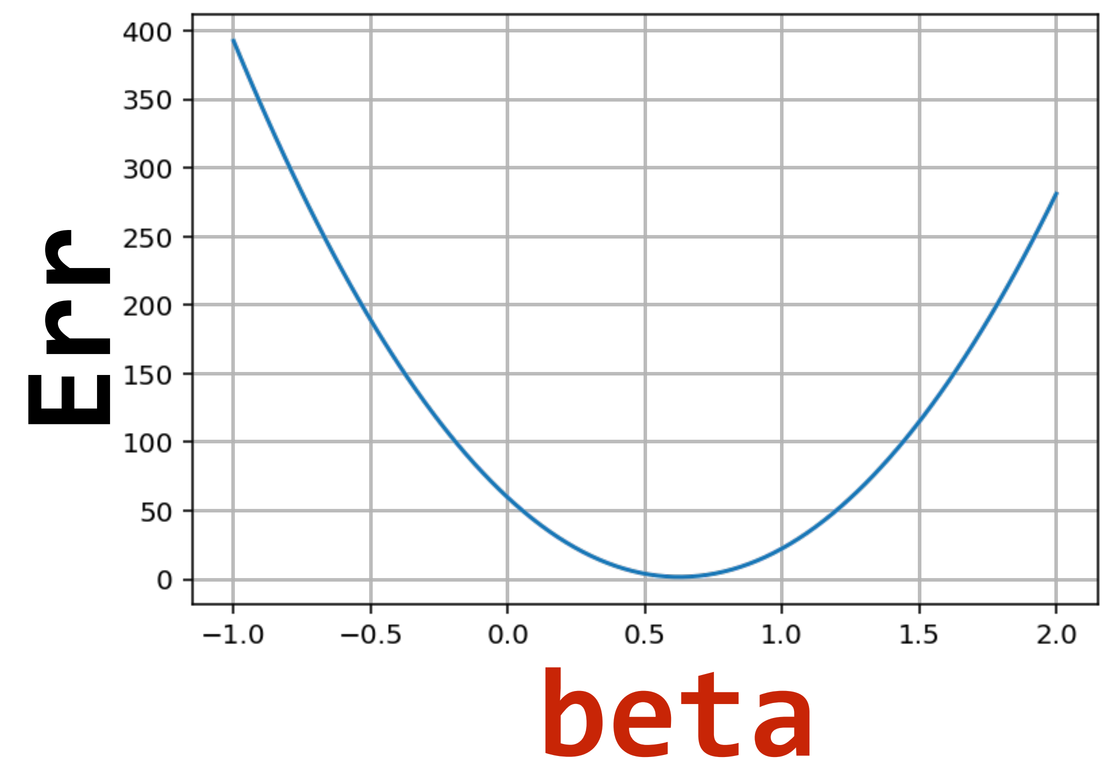


Err(\beta) = \sum_{t=1}^T (y_t - x_t \beta)^2

1D EXAMPLE

$$Err(\beta) = \sum_{t=1}^{T} (y_t - x_t \beta)^2$$

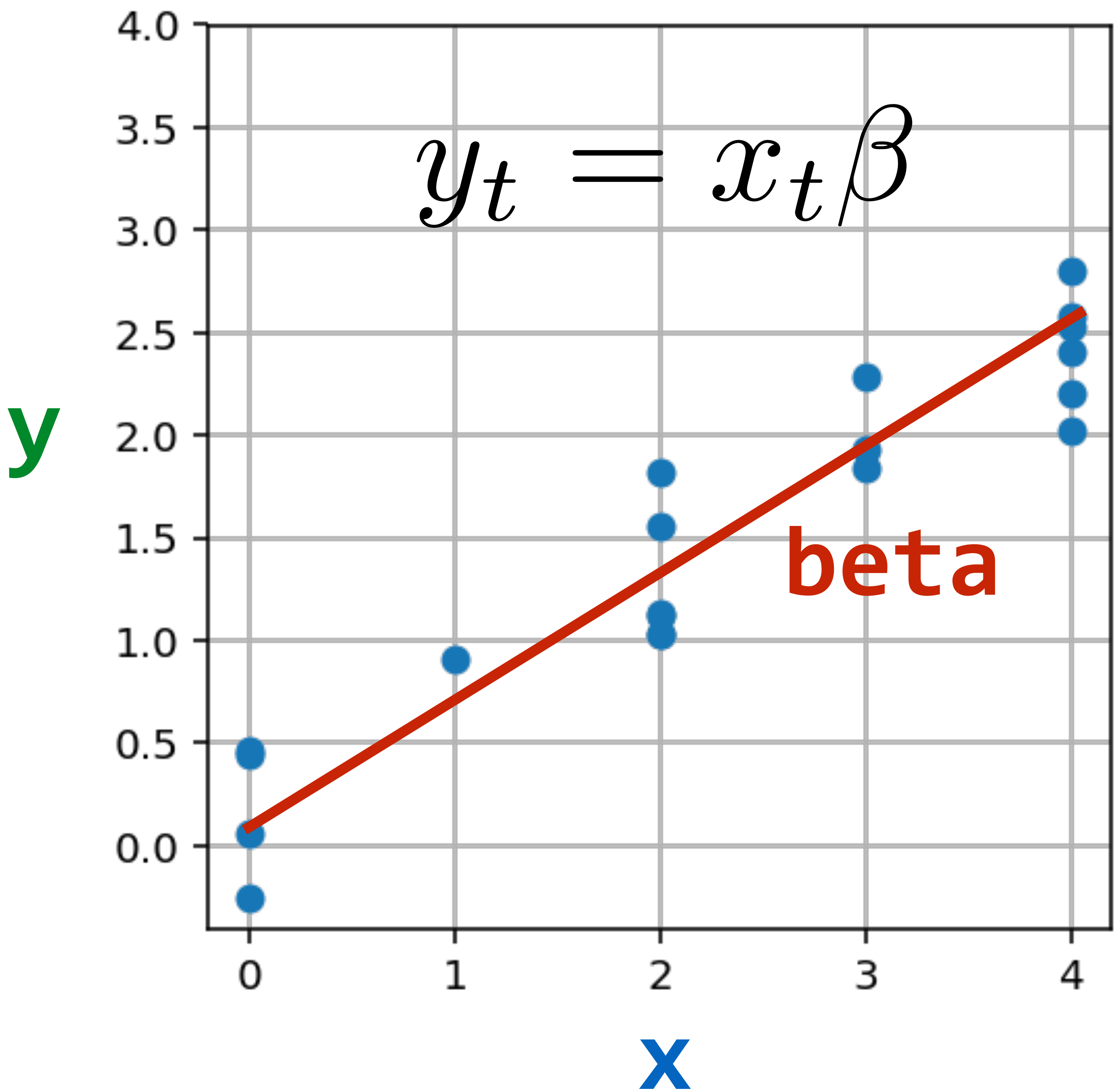


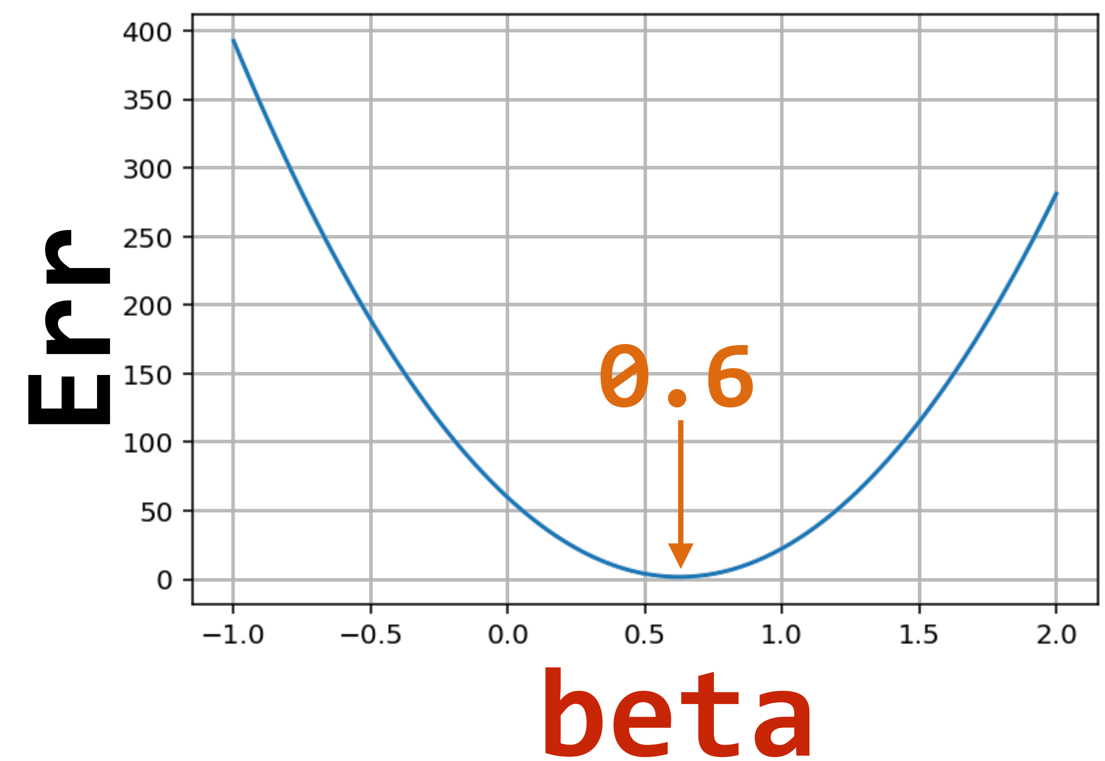


1D EXAMPLE

Err(\beta) = \sum_{t=1}^T (y_t - x_t \beta)^2

$$Err(\beta) = \sum_{t=1}^{T} (y_t - x_t \beta)^2$$

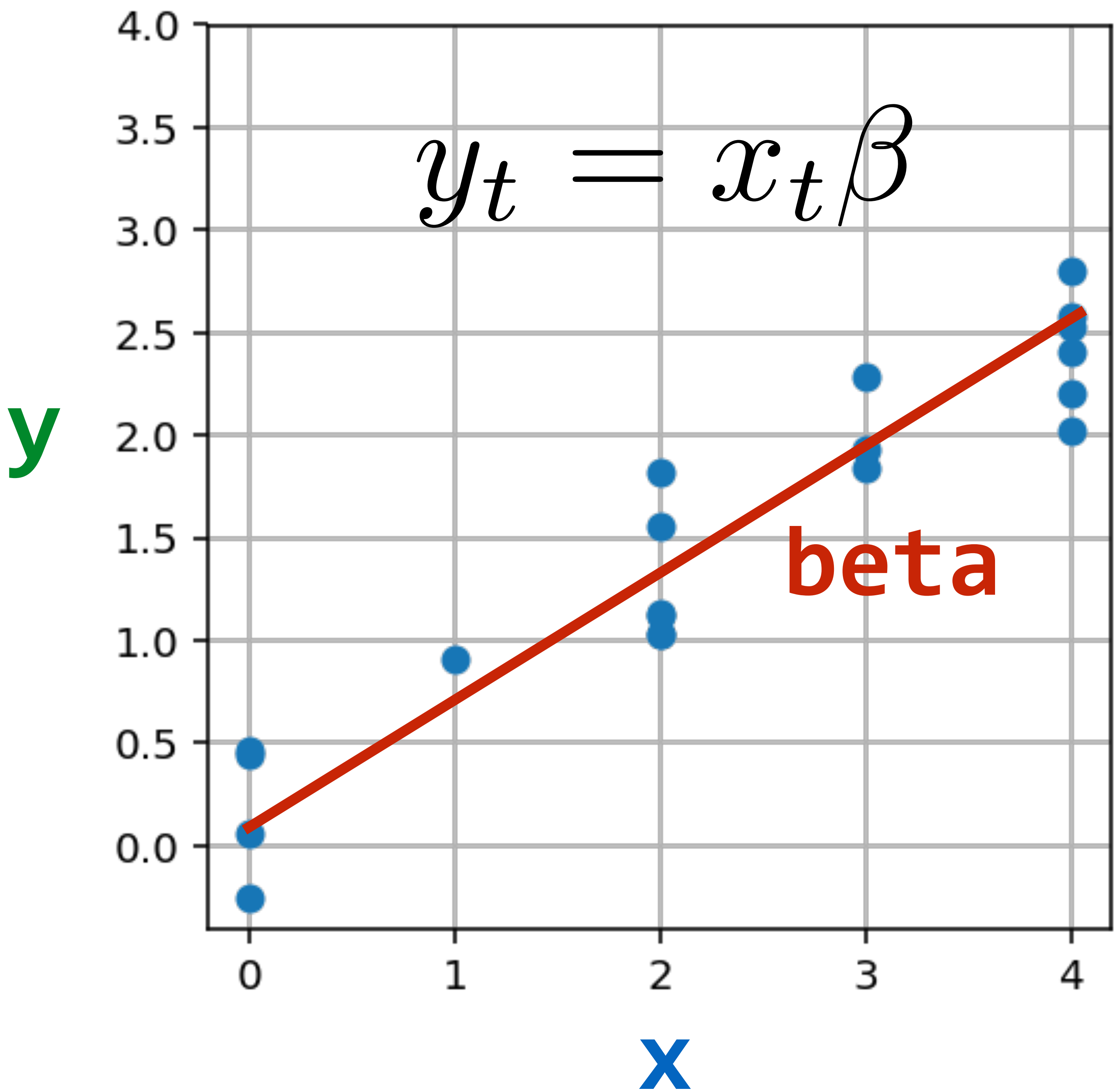


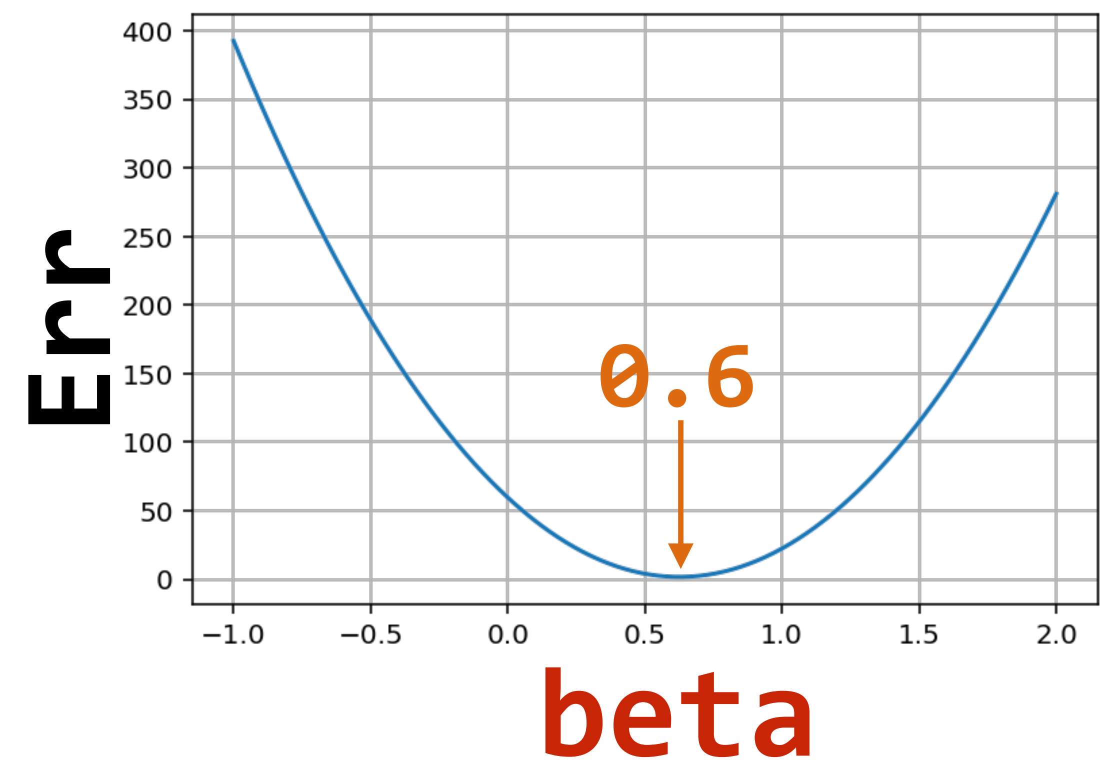


Err(\beta) = \sum_{t=1}^T (y_t - x_t \beta)^2

1D EXAMPLE

$$Err(\beta) = \sum_{t=1}^{T} (y_t - x_t \beta)^2$$



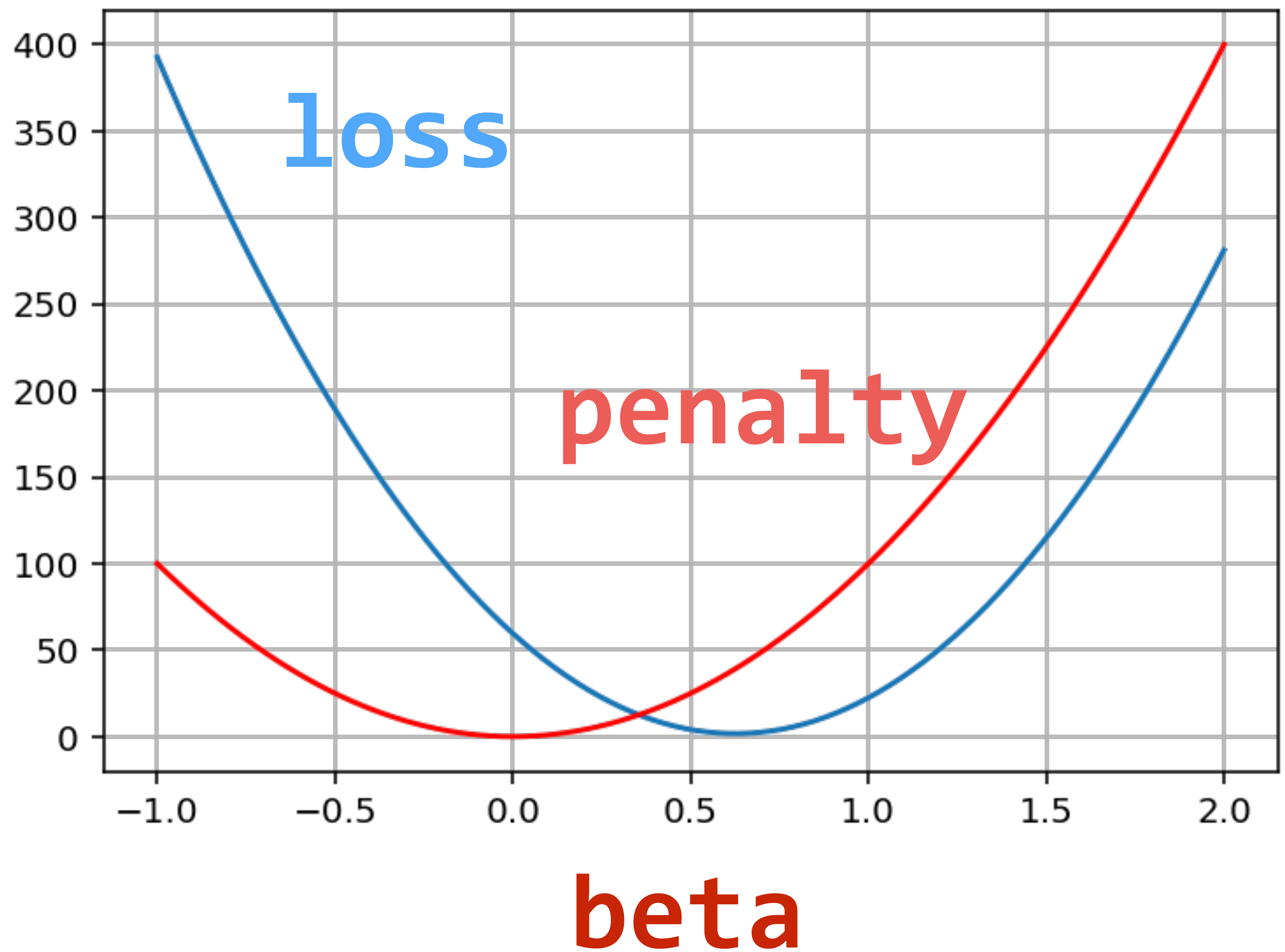


Err(\beta) = \sum_{t=1}^T (y_t - x_t \beta)^2 + \lambda \beta^2

1D EXAMPLE

L2 Regularization:
(as penalty)

$$Err(\beta) = \sum_{t=1}^{T}(y_t - x_t\beta)^2 + \lambda\beta^2$$

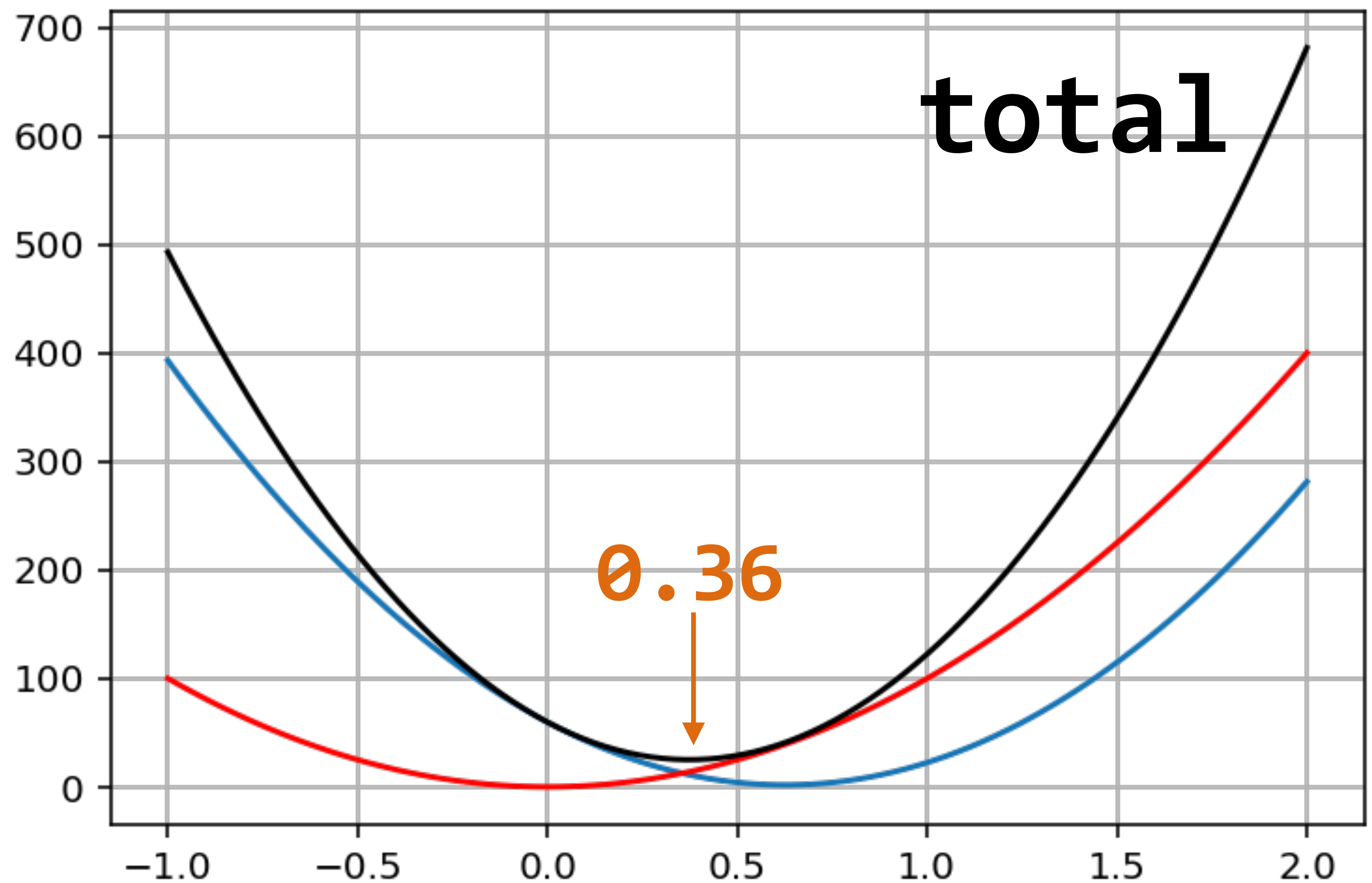


Err(\beta) = \sum_{t=1}^T (y_t - x_t \beta)^2 + \lambda \beta^2

1D EXAMPLE

L2 Regularization:
(as penalty)

$$Err(\beta) = \sum_{t=1}^{T}(y_t - x_t\beta)^2 + \lambda\beta^2$$



beta

1D EXAMPLE

Err(\beta) = \sum_{t=1}^T (y_t - x_t \beta)^2 + \lambda \beta^2

L1 Regularization:
(as penalty)

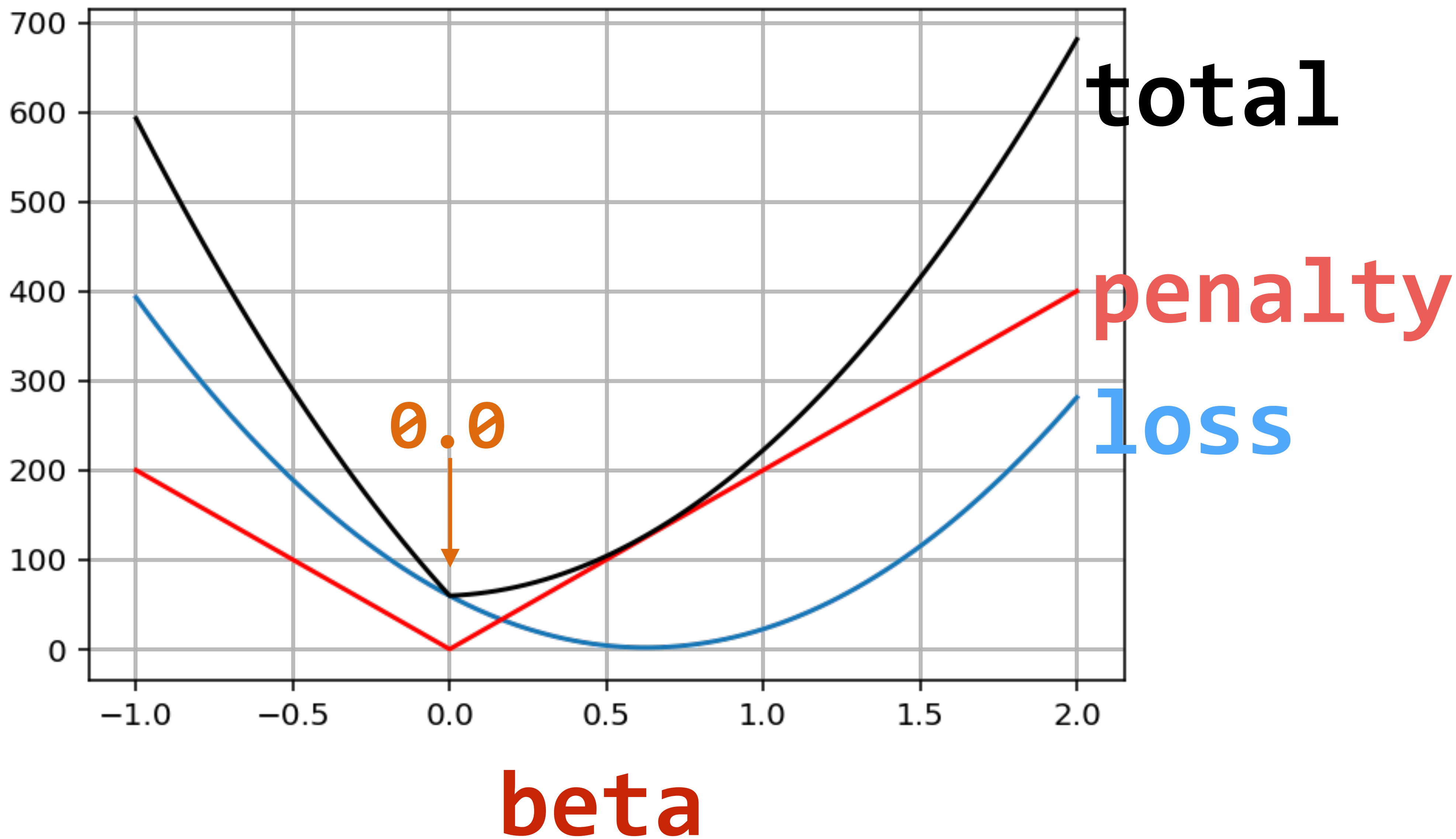
$$Err(\beta) = \boxed{\sum_{t=1}^{T}(y_t - x_t\beta)^2} + \boxed{\lambda|\beta|}$$

1D EXAMPLE

Err(\beta) = \sum_{t=1}^T (y_t - x_t \beta)^2 + \lambda \beta^2

L1 Regularization:
(as penalty)

$$Err(\beta) = \sum_{t=1}^{T}(y_t - x_t\beta)^2 + \lambda|\beta|$$

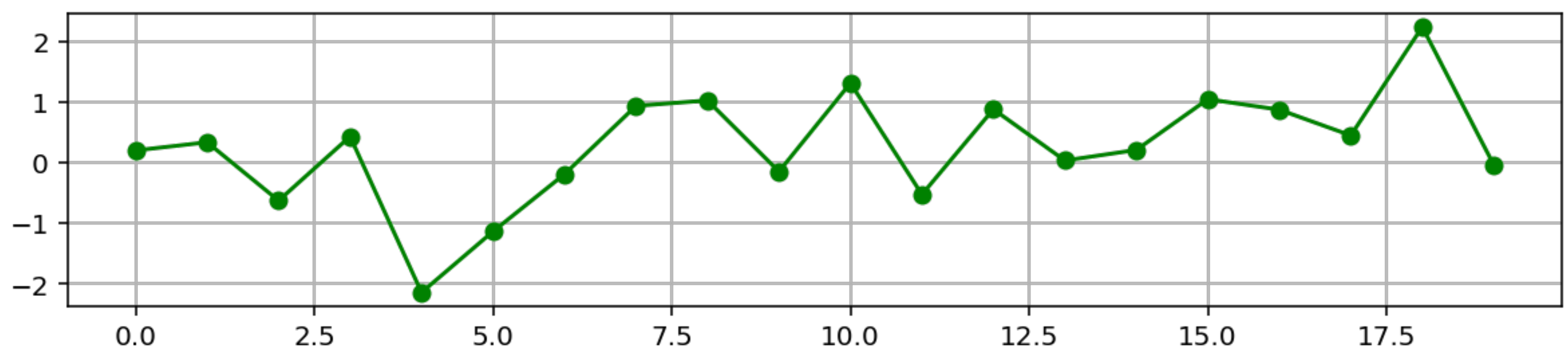


2D EXAMPLE

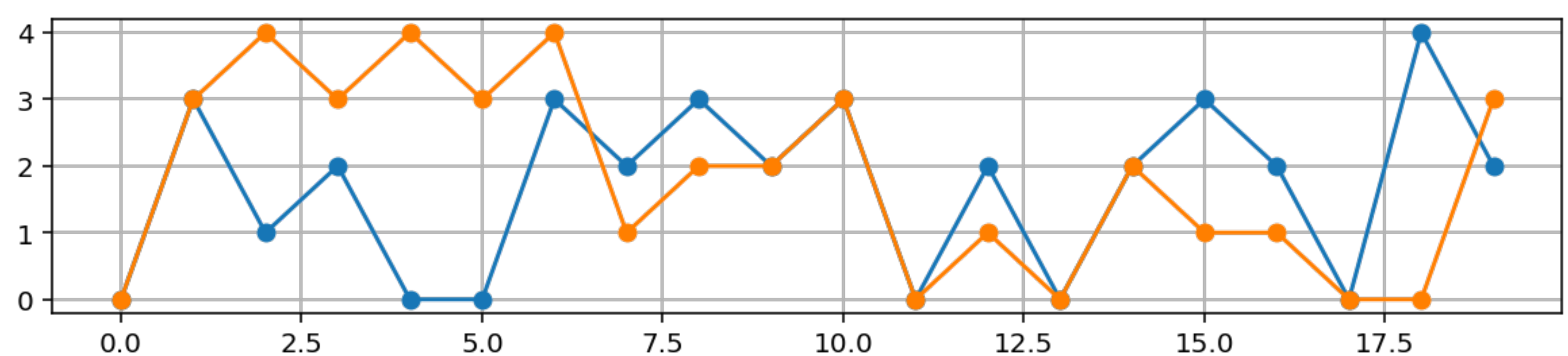
subject

stimulus

y

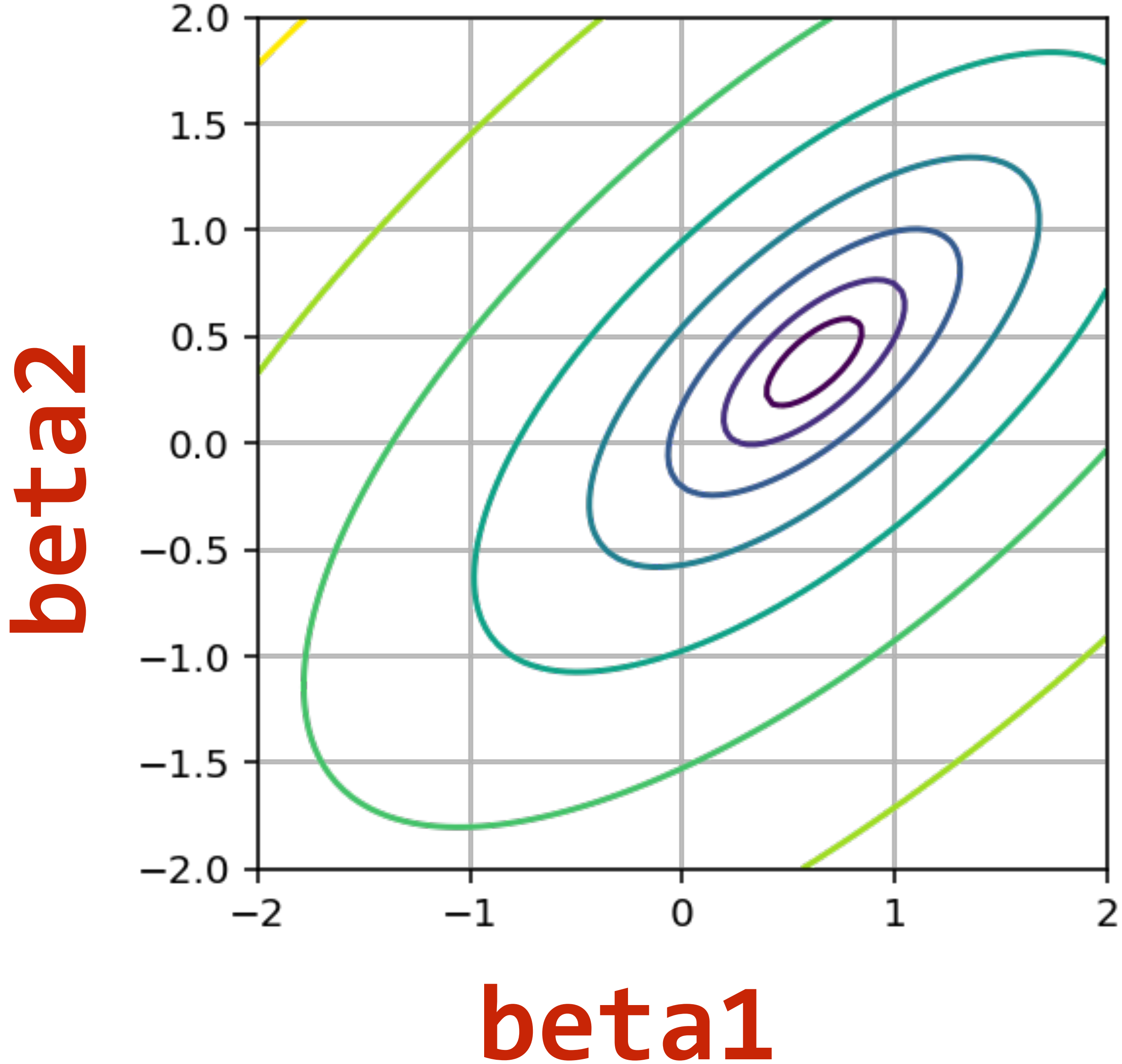


x



2D EXAMPLE

$$y_t = x_{1,t}\beta_1 + x_{2,t}\beta_2$$



LASSO

* Laplacian prior on beta_i

* L1 penalty on beta

* Coordinate descent w/ early stopping

LASSO

\hat\beta = \underset{\beta}{\mbox{argmin}} \left[||Y-X\beta||_2^2 + \lambda ||\beta||_1 \right]

\hat\beta = \underset{\beta}

$$Y = X\beta + \epsilon$$

$$\hat{\beta} = \underset{\beta}{\mathrm{argmin}} \left[||Y - X\beta||_2^2 + \lambda ||\beta||_1 \right]$$

ERROR or LOSS

PENALTY

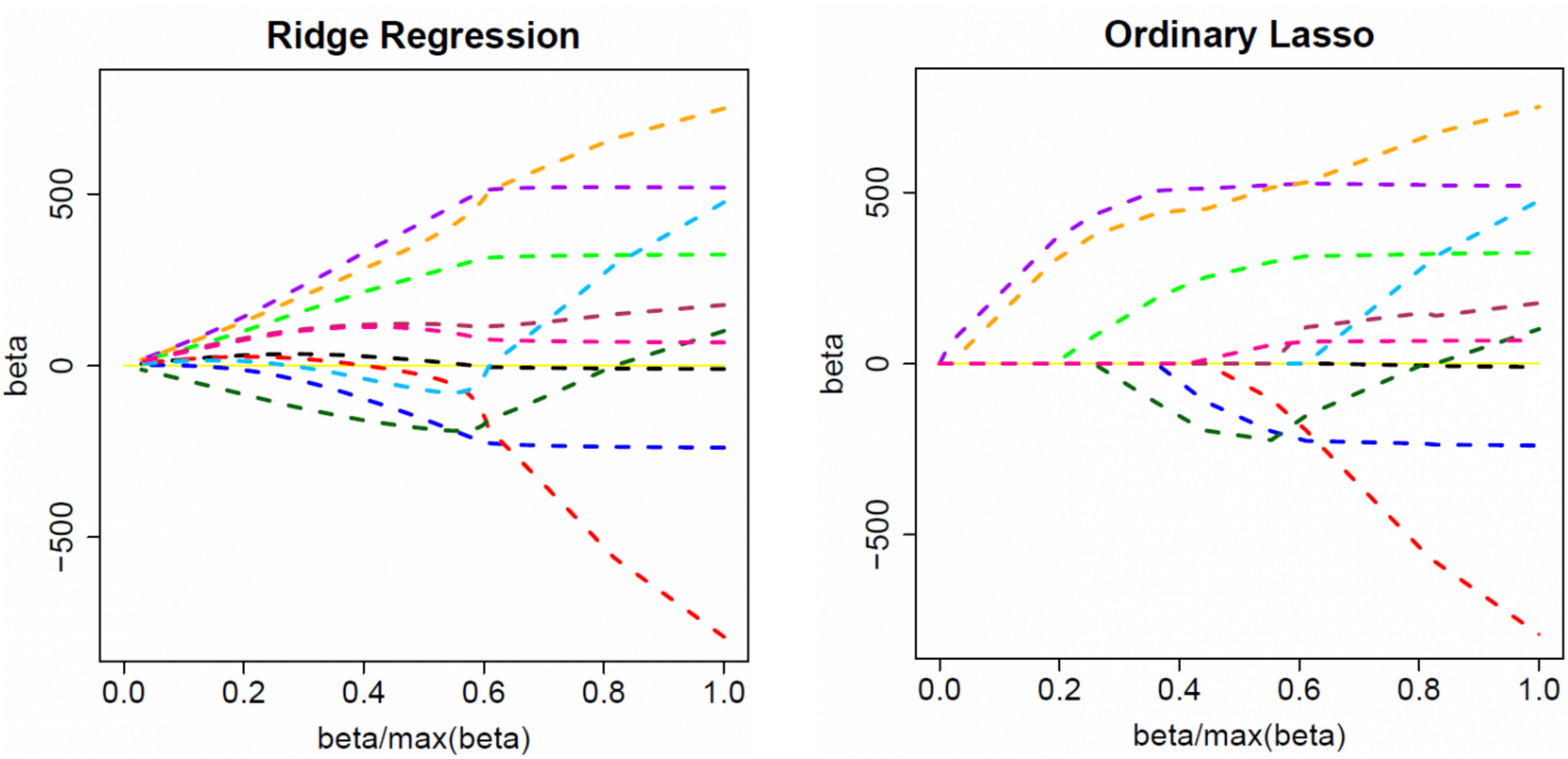
LASSO

* No closed form solution

* Solved via coordinate descent, LARS (least-angle regression) or other methods

* *Sssssllllloooooowwwww……………*

LASSO



h/t https://onlinecourses.science.psu.edu/stat857/node/158

OTHER METHODS

* Neural networks
* Random forests
* Feature selection (~L0-norm)