LINEAR REGRESSION VII

11.18.2020

HOMEWORK 5

* due FRIDAY!!!

RECAP

* np.linalg.lstsq — numpy function that
  does least squares regression (often bad)

* R^2 measures how good a regression model is

* in-set vs. out-of-set model evaluation

* regularized (ridge) regression

RECAP: THE PROBLEM

* Load a dataset containing data from 442 diabetes patients

    * for each patient there are 10 features (e.g. age, sex, bmi, etc.)

    * and 1 outcome ("disease progression after one year")

* We'll be using linear regression to predict disease progression from the 10 features

LINEAR REGRESSION LAB

* If you want to following along, pull the latest version of the **ndap-fa2020** repository from github

  * https://github.com/alexhuth/ndap-fa2020/

* Then see **35-linear_regression-6/35-regression-demos.ipynb**

* Today we're picking up where we left off on Monday (see **36-linear_regression-7**)

WHY IS RIDGE REGRESSION CHANGING THE PATTERN OF WEIGHTS INSTEAD OF JUST MAKING THE WEIGHTS SMALLER?
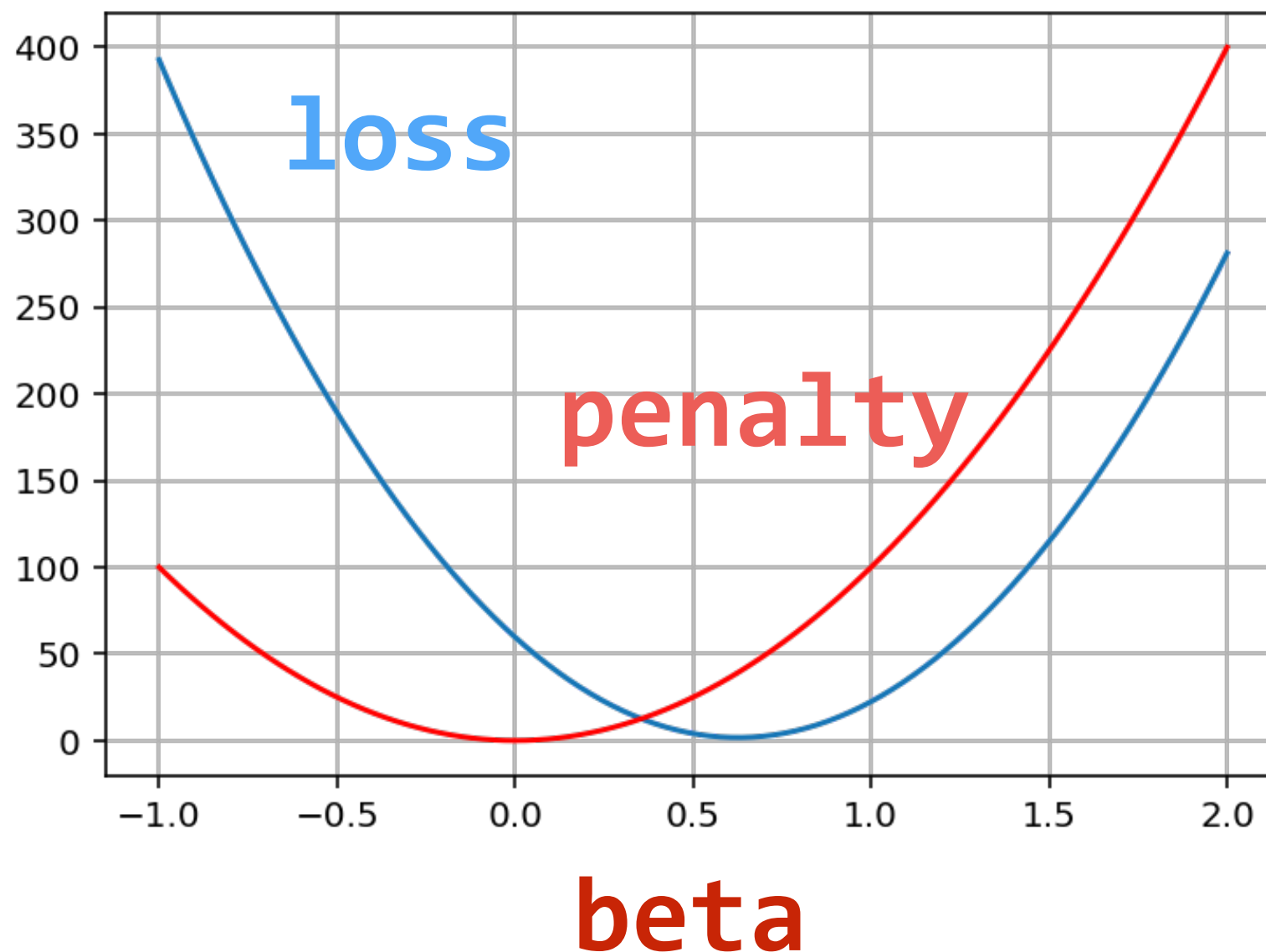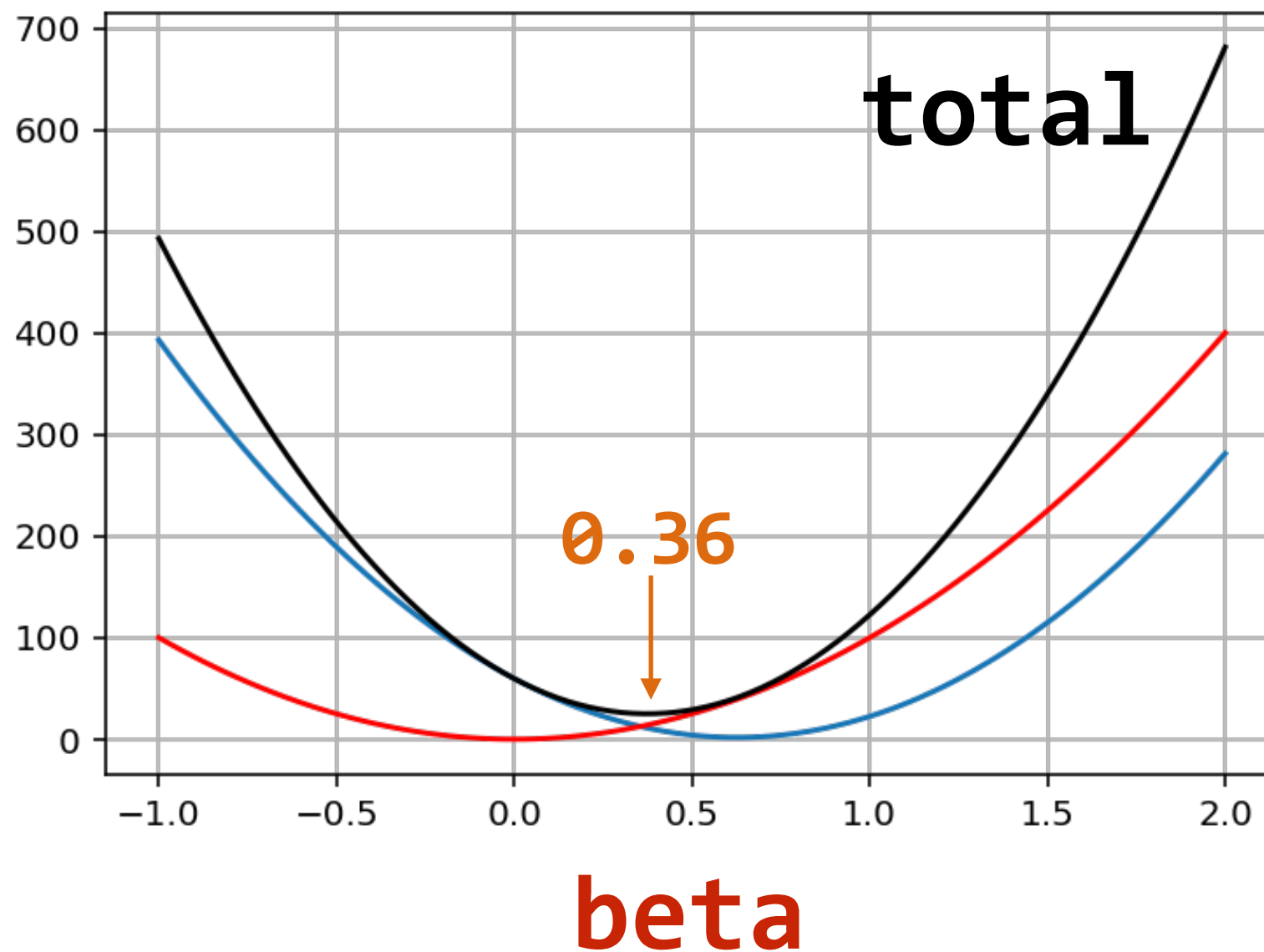
1D EXAMPLE
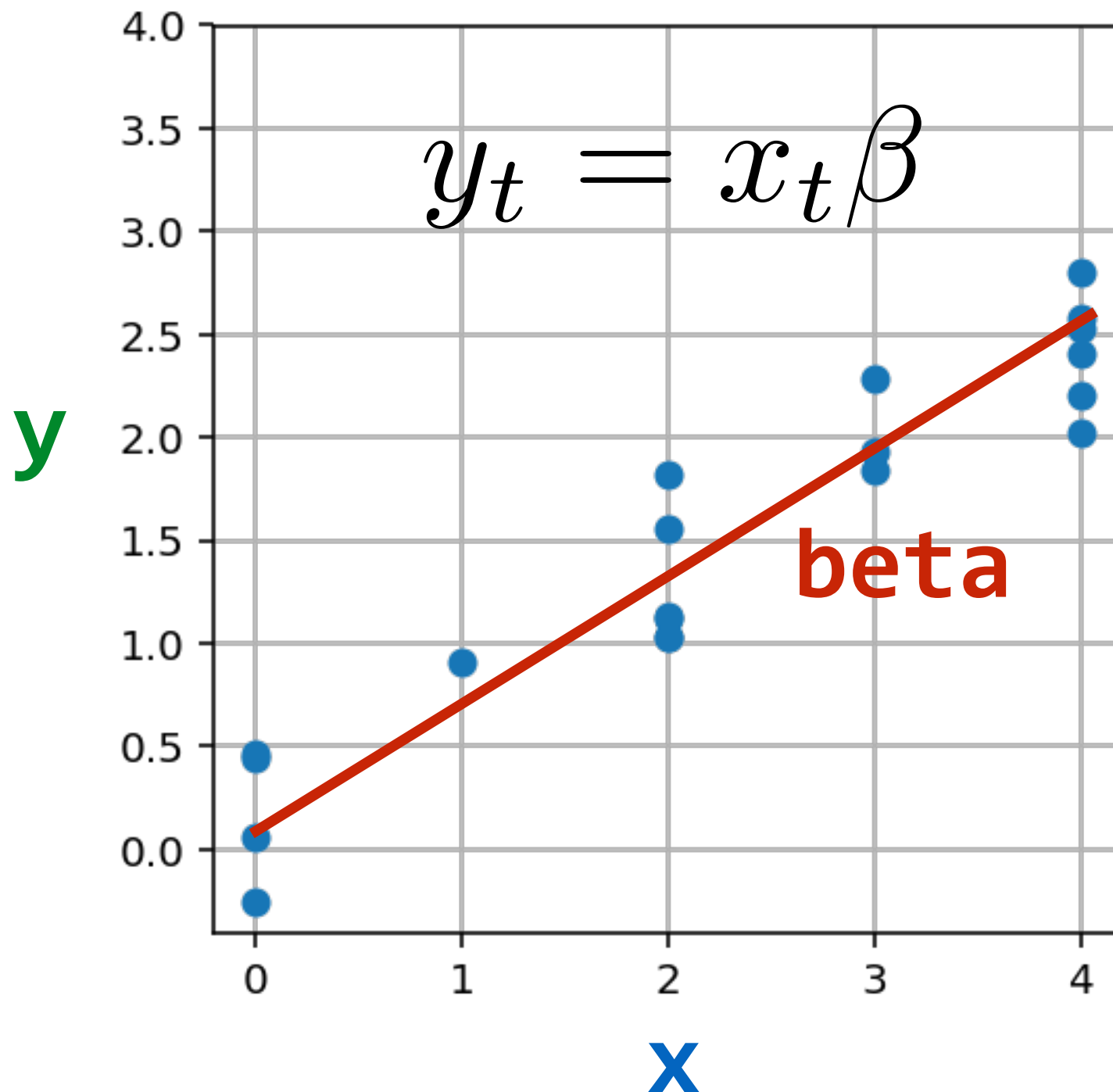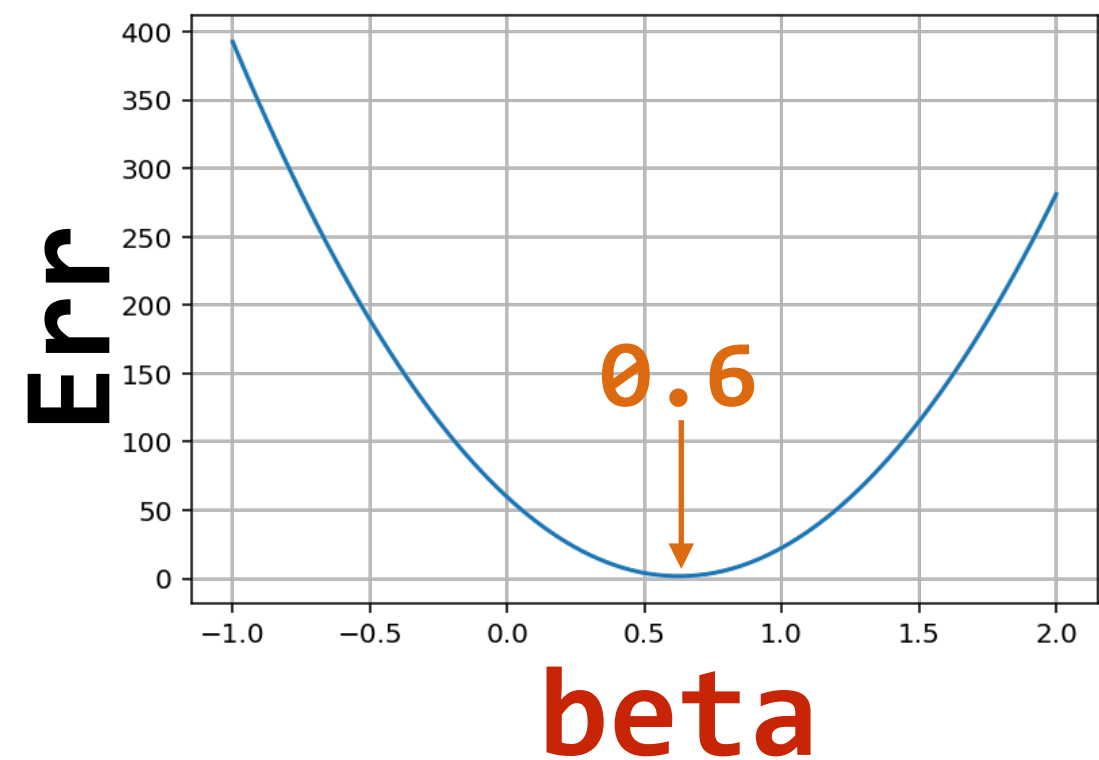
$$y_t = x_t \beta$$

beta

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

0.6

beta

1D EXAMPLE

L2 Regularization: (as penalty)

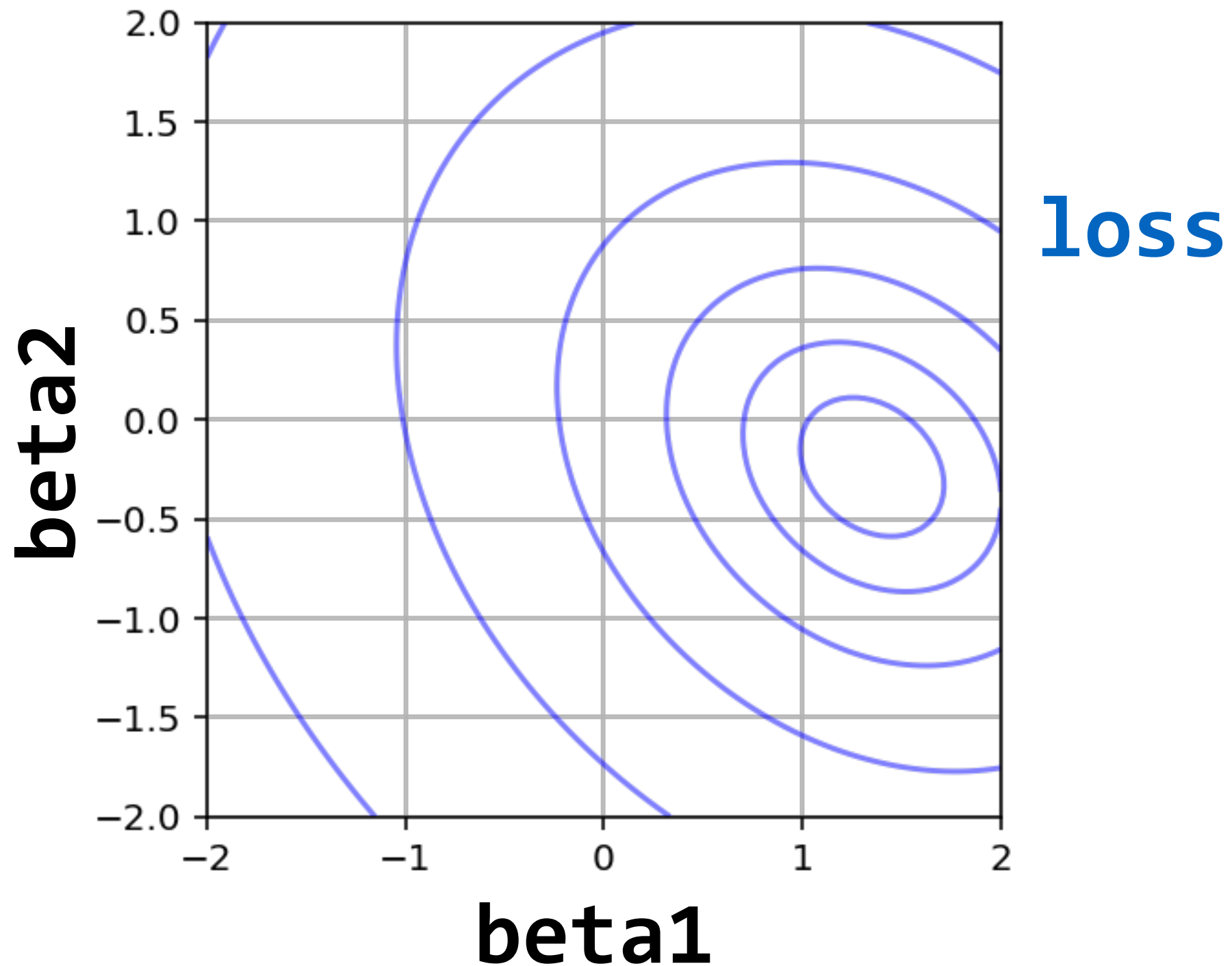$$Err(\beta) = \sum_{t=1}^{T}(y_t - x_t\beta)^2 + \lambda\beta^2$$

loss

penalty

beta

1D EXAMPLE

L2 Regularization:
(as penalty)

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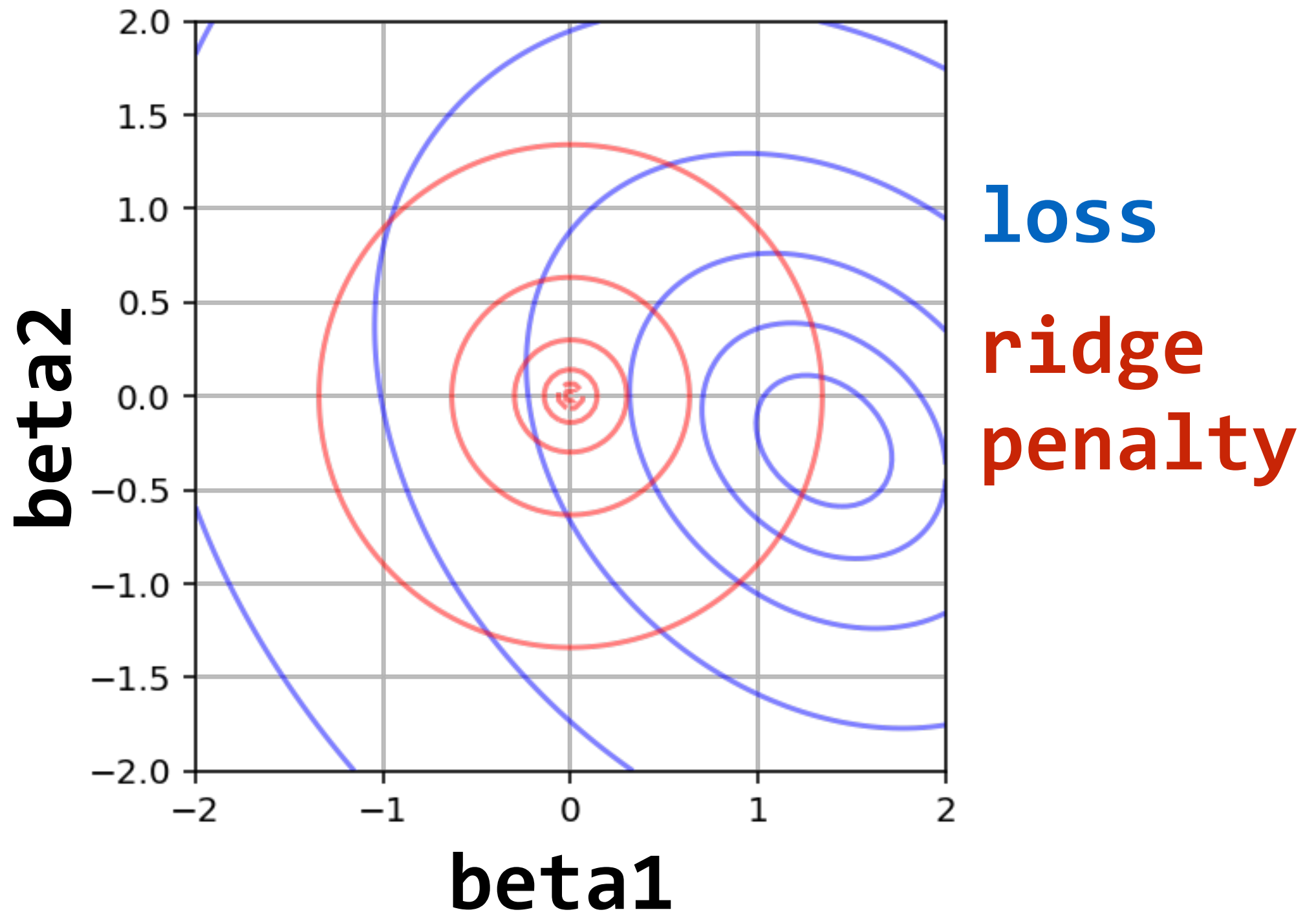

total

0.36

beta

2D EXAMPLE

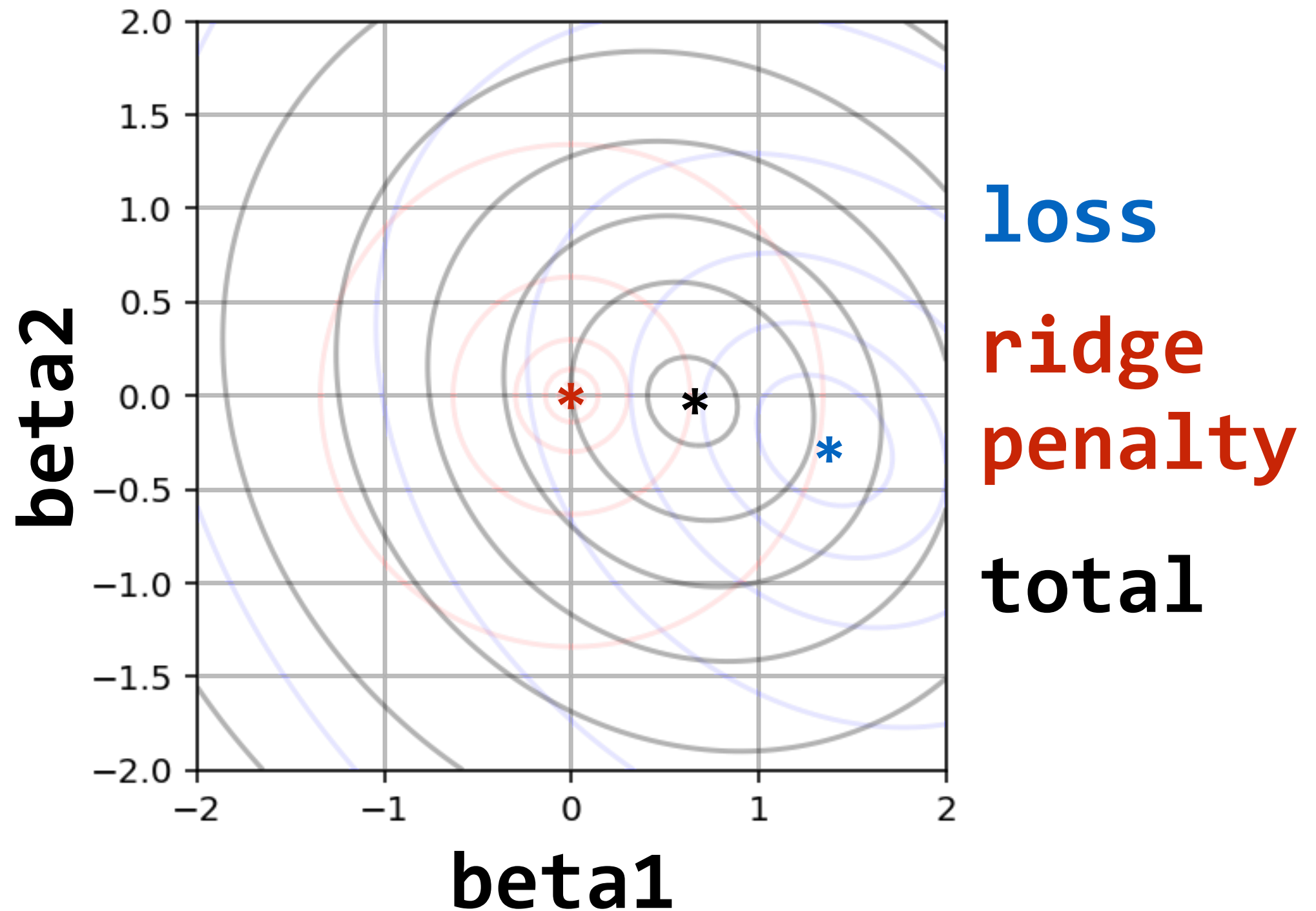$$y_t = x_{1,t}\beta_1 + x_{2,t}\beta_2$$

loss

2D EXAMPLE

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

2D EXAMPLE

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

REGRESSION TAKEAWAYS

* You want to predict one thing (y) from other things (x1, x2, etc.)? **Regression!**

* Regression works best if it's **regularized**

* You've seen **ridge regression,** but there are many others!

  * Tikhonov regression, sparse regression (LASSO), elastic net regression, etc.

END