MODEL FITTING & TIKHONOV REGRESSION

Prof. Alexander Huth

9/26/2017

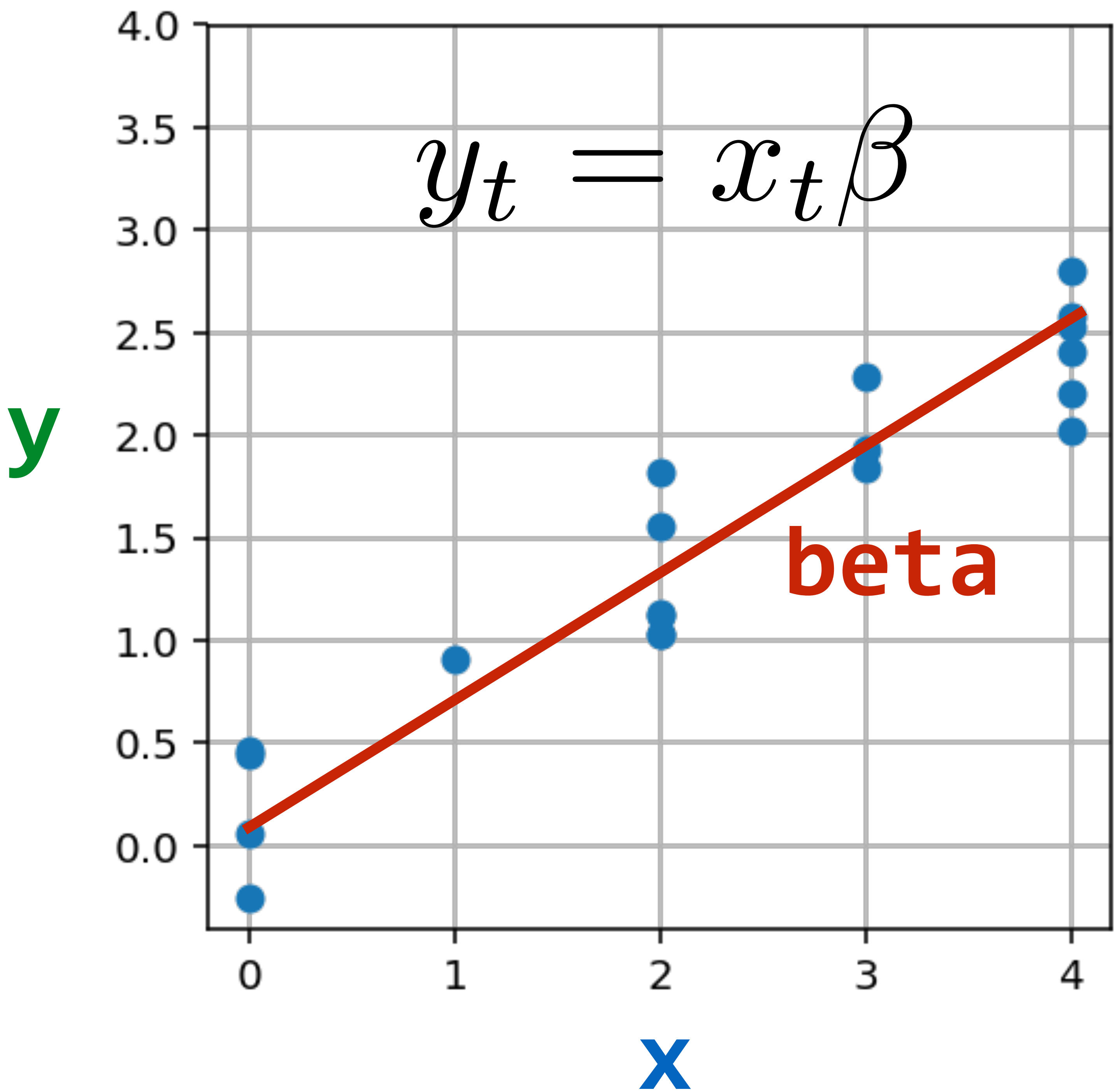
RECAP

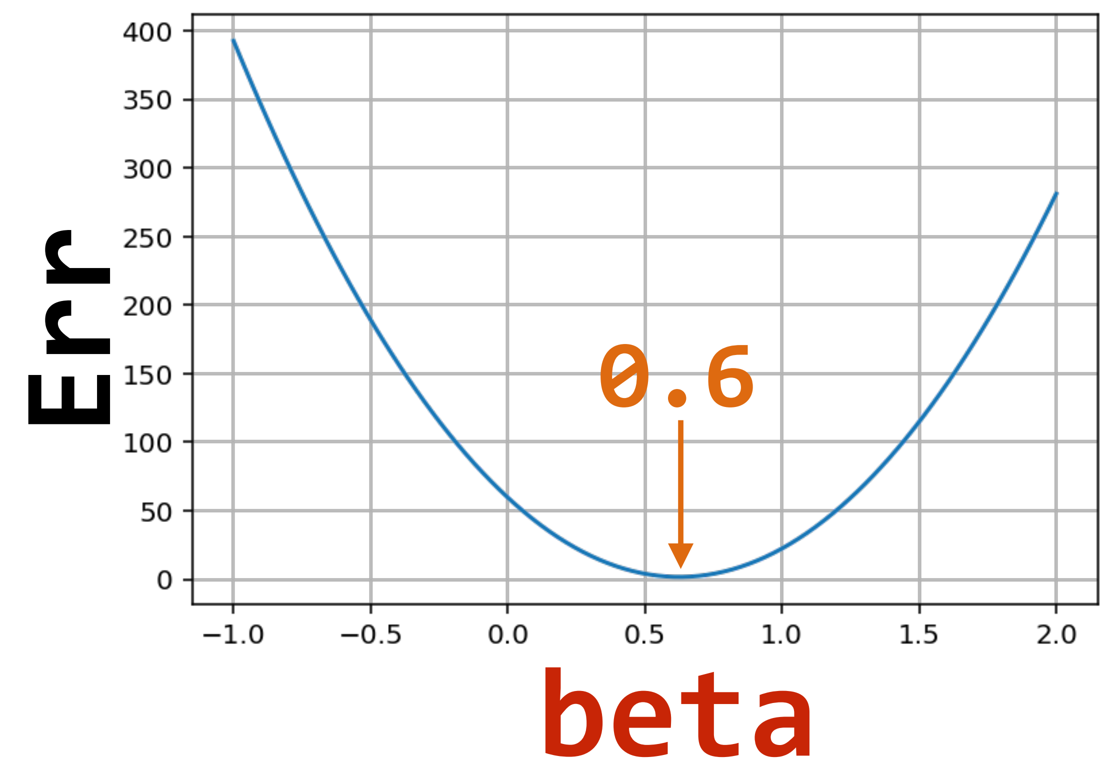
* Regularized linear regression!!!

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

1D EXAMPLE

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



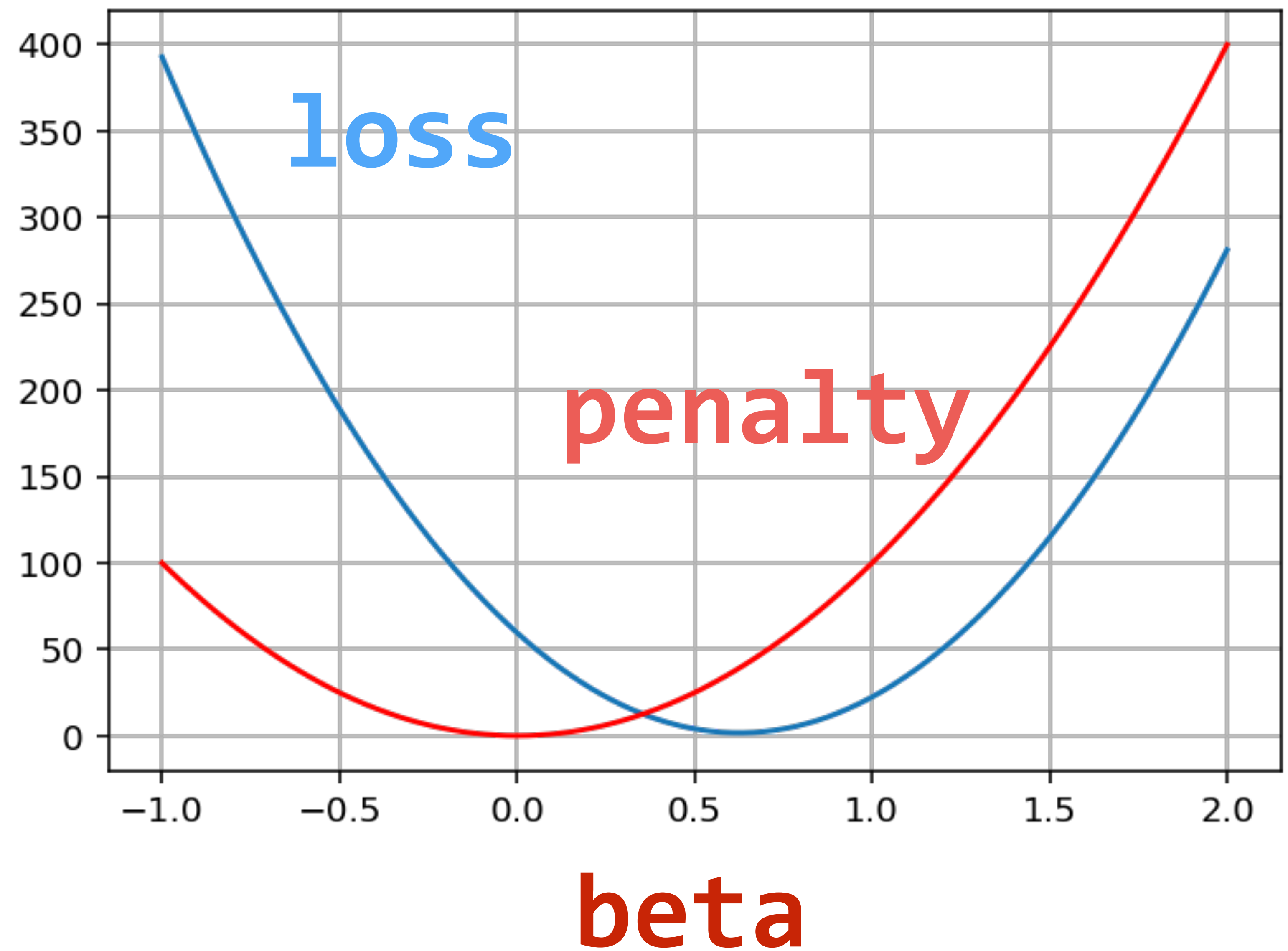


Err(\beta) = \sum_{t=1}^T (y_t - x_t \beta)^2 + \lambda \beta^2

1D EXAMPLE

L2 Regularization:
(as penalty)

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

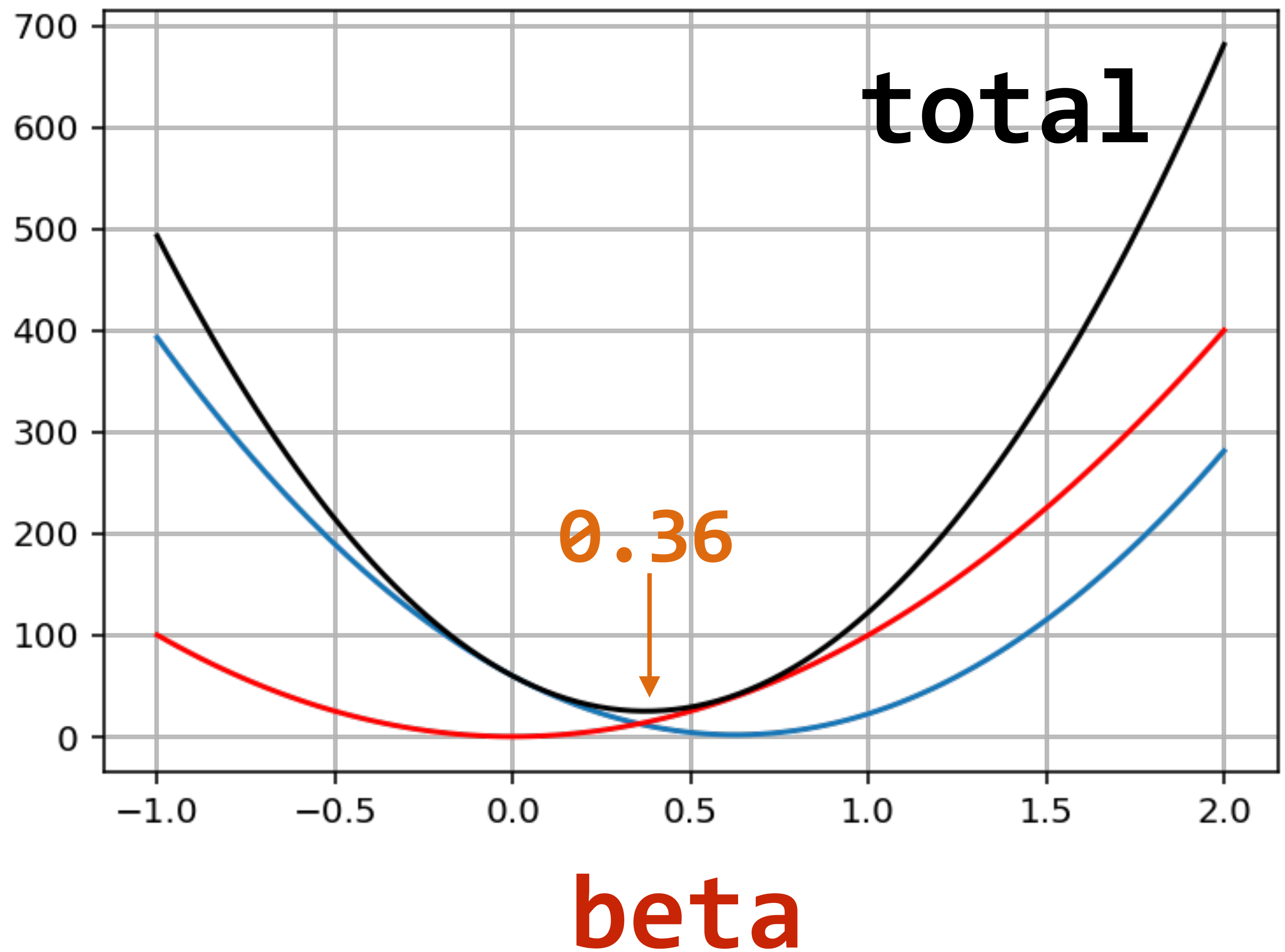


1D EXAMPLE

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

L2 Regularization:
(as penalty)

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



RIDGE REGRESSION

* Multivariate normal (MVN) prior on beta

* L2 penalty on beta

* Gradient descent w/ early stopping

RIDGE REGRESSI

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

\hat\beta = \underset{\beta}

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

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

ERROR or LOSS

PENALTY

RIDGE REGRESSION

$$\hat\beta = (X^\top X + \lambda I)^{-1} X^\top Y$$

$$\hat\beta = X^+_{ridge} Y$$

\hat\beta = (X^\top X + \lambda I)^{-1}
X^\top Y

RIDGE REGRESSION

* Efficient solution with SVD

$$\hat{\beta} = (X^\top X + \lambda I)^{-1} X^\top Y$$

(SVD) $X = USV^\top$ $\quad D = \dfrac{S}{S^2 + \lambda^2}$

$$\hat{\beta} = V D U^\top Y$$

\hat\beta = (X^\top X + \lambda I)^{-1} X^\top Y

D = \frac{S}{S^2 + \lambda^2}
\hat\beta = V D U^\top Y

COVARIANCE

* A measure of the “joint variability” of two variables

COVARIANCE

\mbox{cov}(\vec{x},\vec{y}) = \frac{1}{n
\sum_{i=1}^n(x_i-\bar{x})(y_i-\bar{y})

\mbox{cov}(X,Y) = E[(X-E[X])(Y-E[Y])]

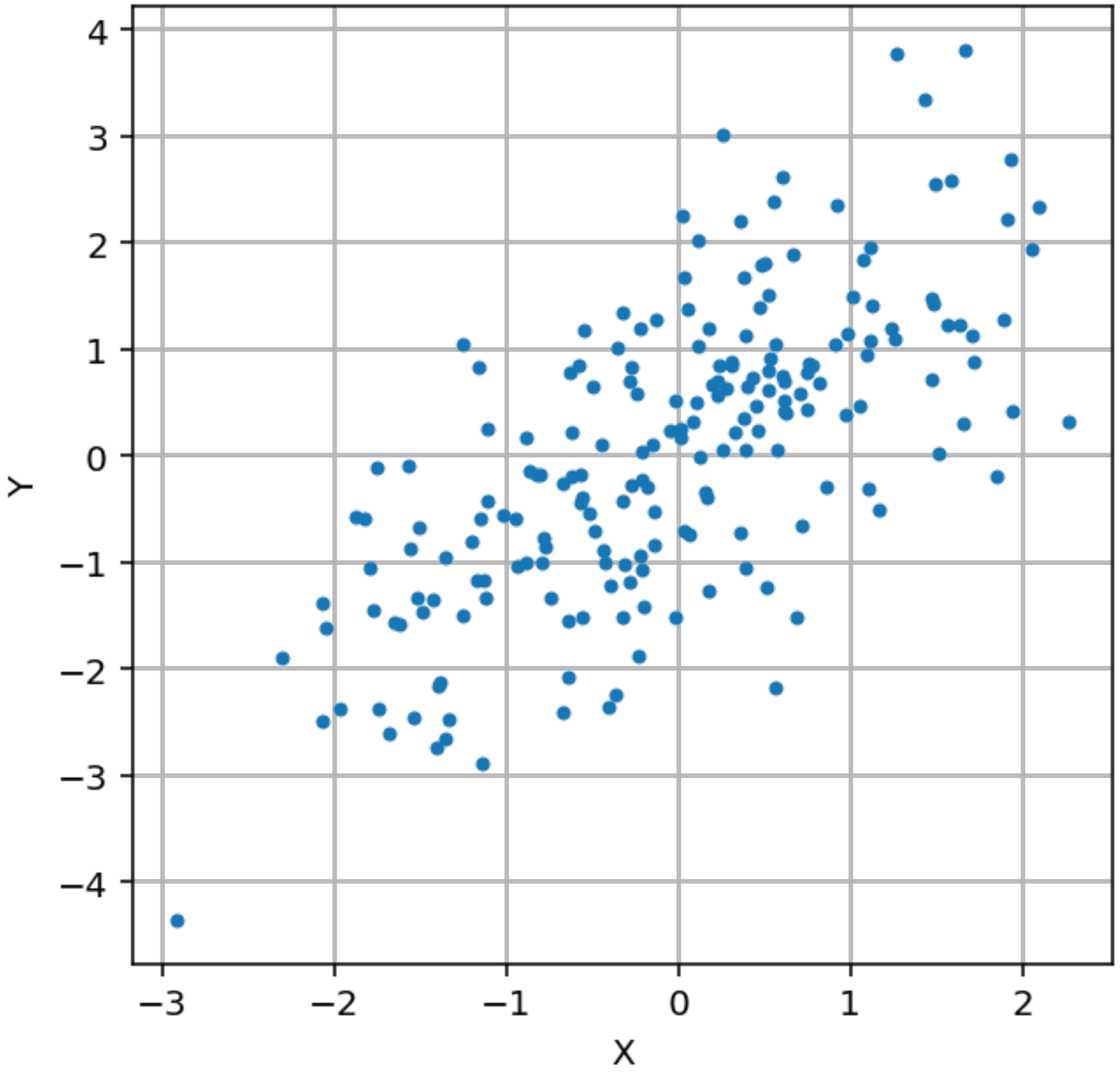
\mbox{cov}(X,X) = E[(X-E[X])(X-E[X])] =
E[(X-E[X])^2] = \mbox{var}(X)

$$\mbox{cov}(\vec{x},\vec{y}) = \frac{1}{n}\sum_{i=1}^n(x_i-\bar{x})(y_i-\bar{y})$$

$$\mbox{cov}(X,Y) = E[(X-E[X])(Y-E[Y])]$$

$$\mbox{cov}(X,X) = E[(X-E[X])(X-E[X])] = E[(X-E[X])^2] = \mbox{var}(X)$$

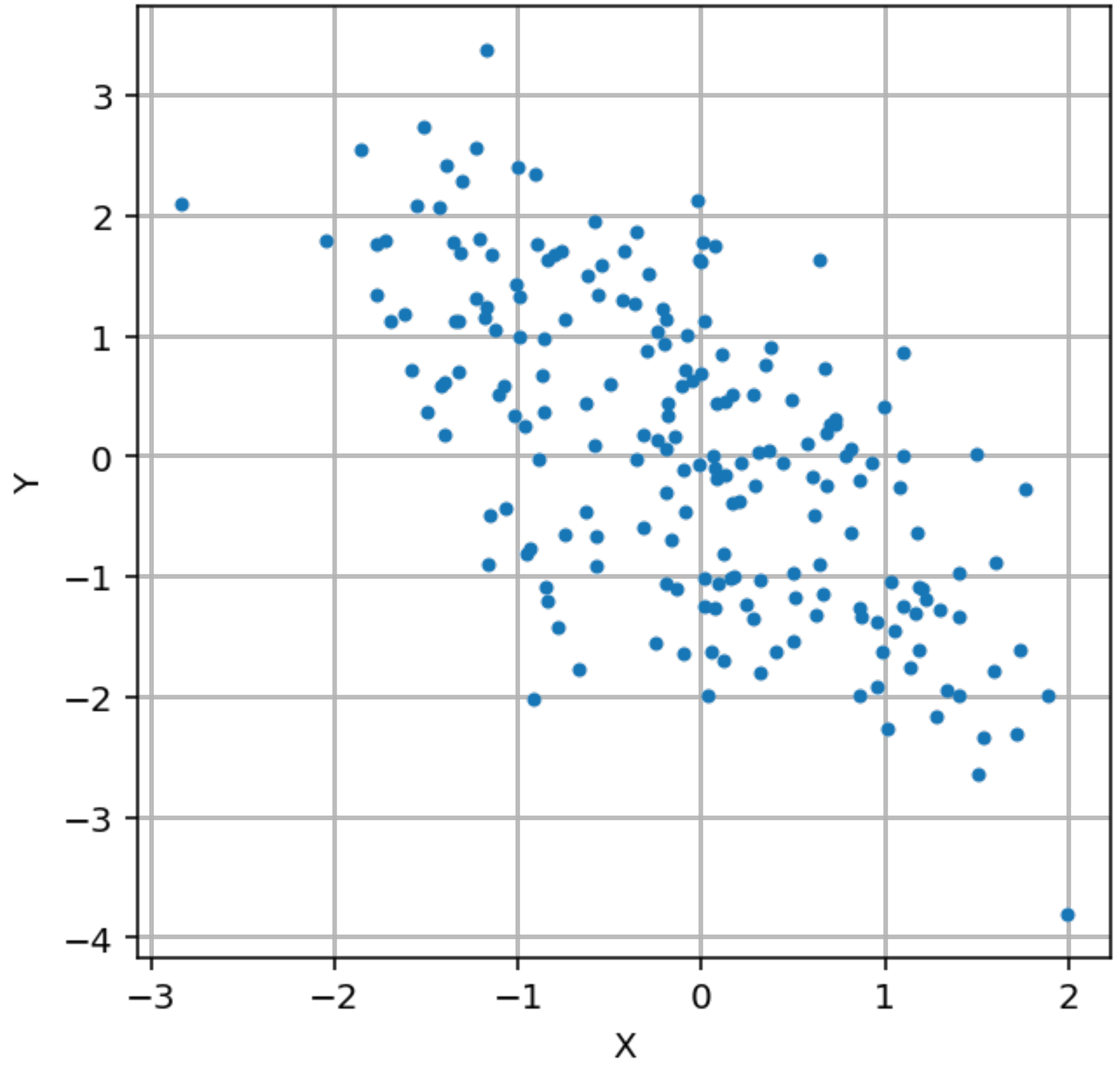
COVARIANCE



cov(x,y) > 0?

cov(x,y) < 0?

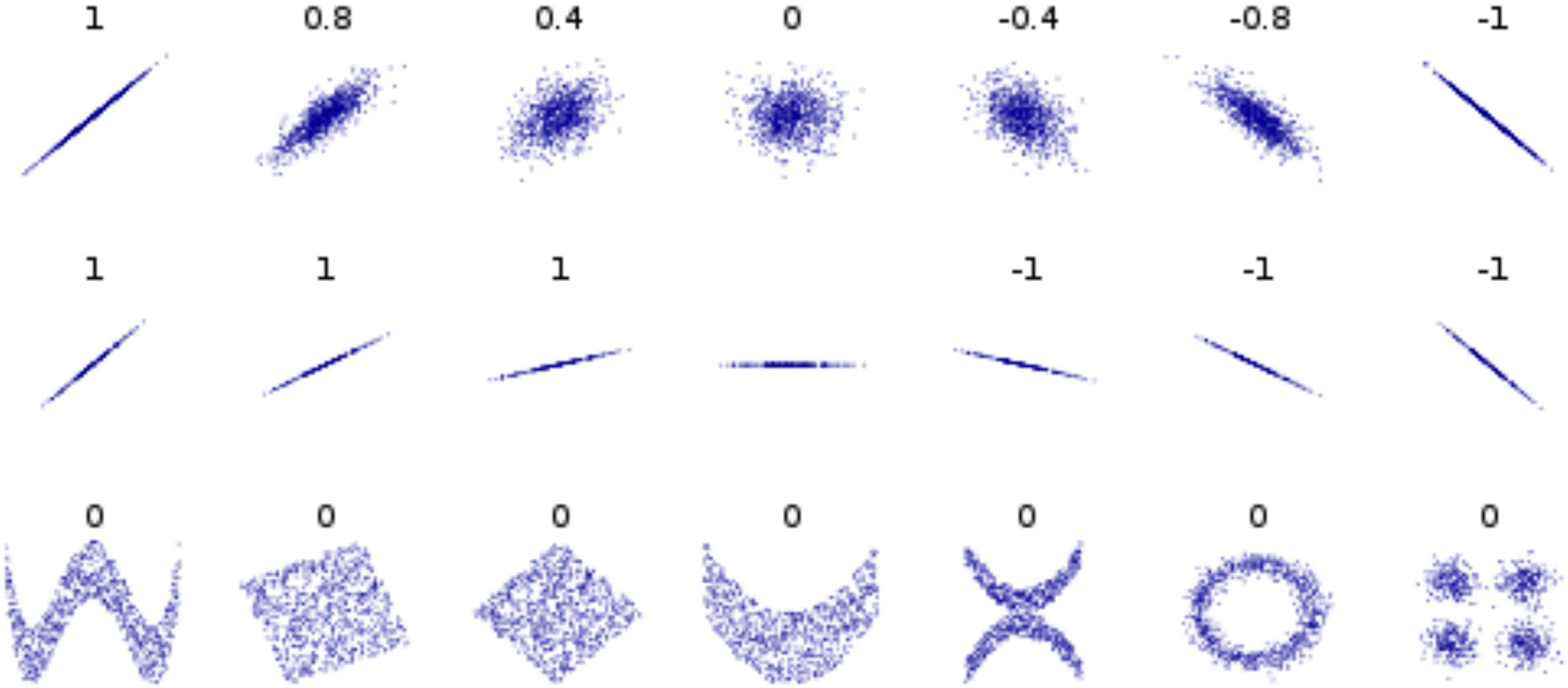
COVARIANCE



cov(x,y) > 0?

cov(x,y) < 0?

COVARIANCE



COVARIANCE MATRIX

\mbox{cov}(A) = \begin{bmatrix}
\mbox{var}(A_1) & \mbox{cov}(A_
\\ \mbox{cov}(A_1,A_2) & \mbox{v
(A_2) \end{bmatrix}

\mbox{cov}(A) = \left(\frac{1}{n}\ri
A^T

$$\begin{array}{cc} & n \\ A = & \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} \end{array} \; 2$$

$$\operatorname{cov}(A) = \begin{bmatrix} \operatorname{var}(A_1) & \operatorname{cov}(A_1, A_2) \\ \operatorname{cov}(A_1, A_2) & \operatorname{var}(A_2) \end{bmatrix}$$

(assuming A is mean 0)

$$\operatorname{cov}(A) = \left(\frac{1}{n}\right) AA^T$$

TIKHONOV REGRESSION

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

\hat\beta = \underset{\beta}

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

* RIDGE REGRESSION

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

ERROR or LOSS PENALTY

* TIKHONOV REGRESSION

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

PENALTY MATRIX

TIKHONOV REGRESSION

* RIDGE REGRESSION is a special case of TIKHONOV REGRESSION

* TIKHONOV REGRESSION puts a **ZERO-MEAN MULTIVARIATE NORMAL PRIOR** on the weights

* in RIDGE REGRESSION the covariance matrix of the prior has a constant diagonal

  * i.e. the prior is a **SPHERE**

* in TIKHONOV REGRESSION the covariance matrix can be ***ANYTHING***

TIKHONOV REGRESSION

* the multivariate normal prior given by **TIKHONOV REGRESSION**

$$\beta \sim N(0, \sigma^2 (C^T C)^{-1})$$

\beta \sim N(0, \sigma^2
\Lambda^{-1}), \Lambda = C^T C

TIKHONOV REGRESSION

```
\begin{eqnarray*}
A &=& XC^{-1}\\
\hat\beta_A &=& \unc
{\mbox{argmax}} \left
\lambda ||\beta||_2^2
```

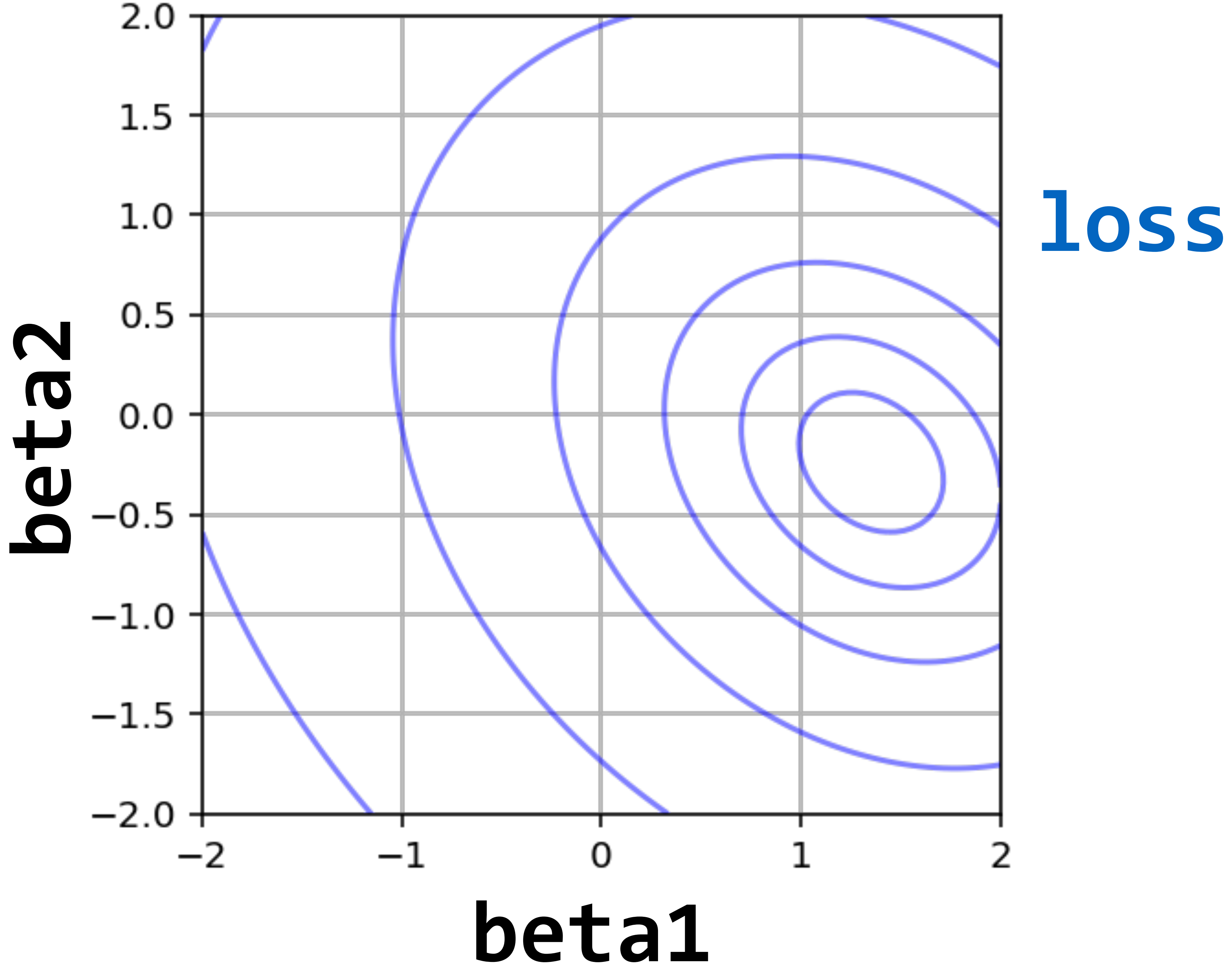
* any **TIKHONOV** problem can be converted into a **RIDGE** problem

$$\begin{aligned}
A &= XC^{-1} && \leftarrow \text{1. CHANGE OF BASIS}\\
\hat{\beta}_A &= \underset{\beta}{\text{argmin}}\left[||Y - A\beta||_2^2 + \lambda||\beta||_2^2\right] && \\
&\qquad\qquad\uparrow \text{ 2. RIDGE REGRESSION} && \\
\hat{\beta} &= C^{-1}\hat{\beta}_A && \leftarrow \text{3. CHANGE BASIS AGAIN}
\end{aligned}$$

Rank-Deficient and Discrete Ill-Posed Problem: Numerical Aspects of Linear Inversion (1998; Per Christian Hansen)

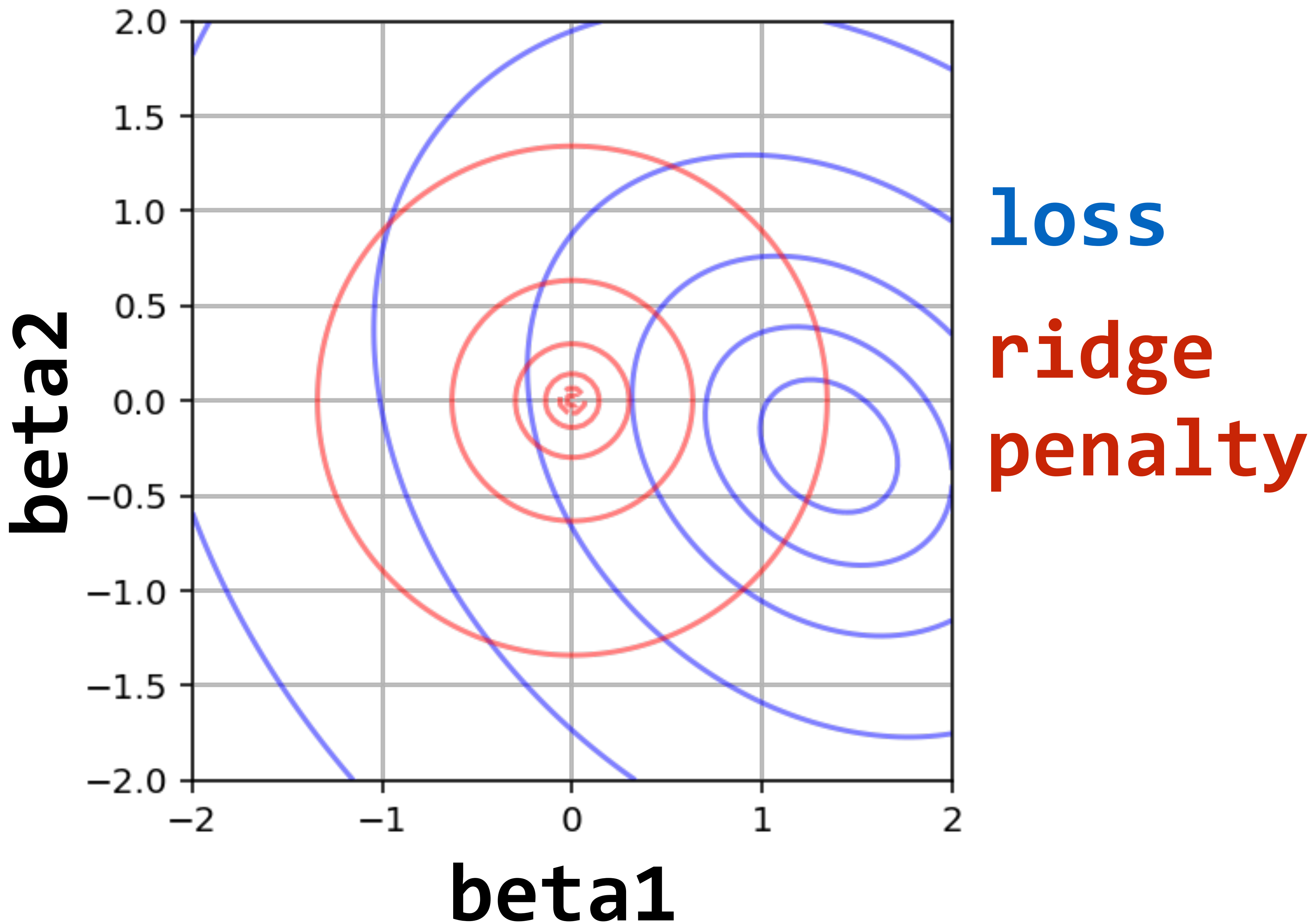
EXAMPLE

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



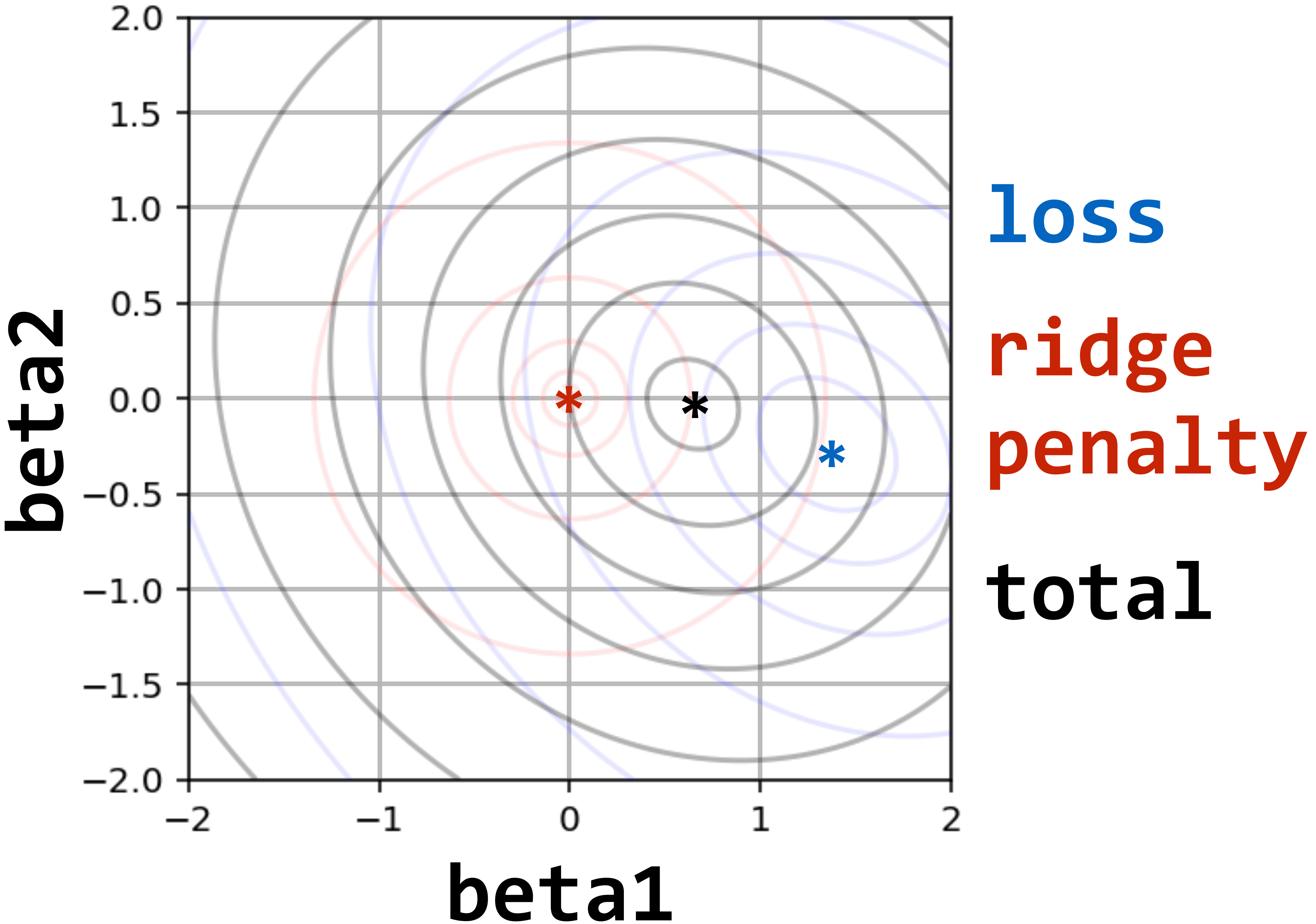
EXAMPLE

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



EXAMPLE

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

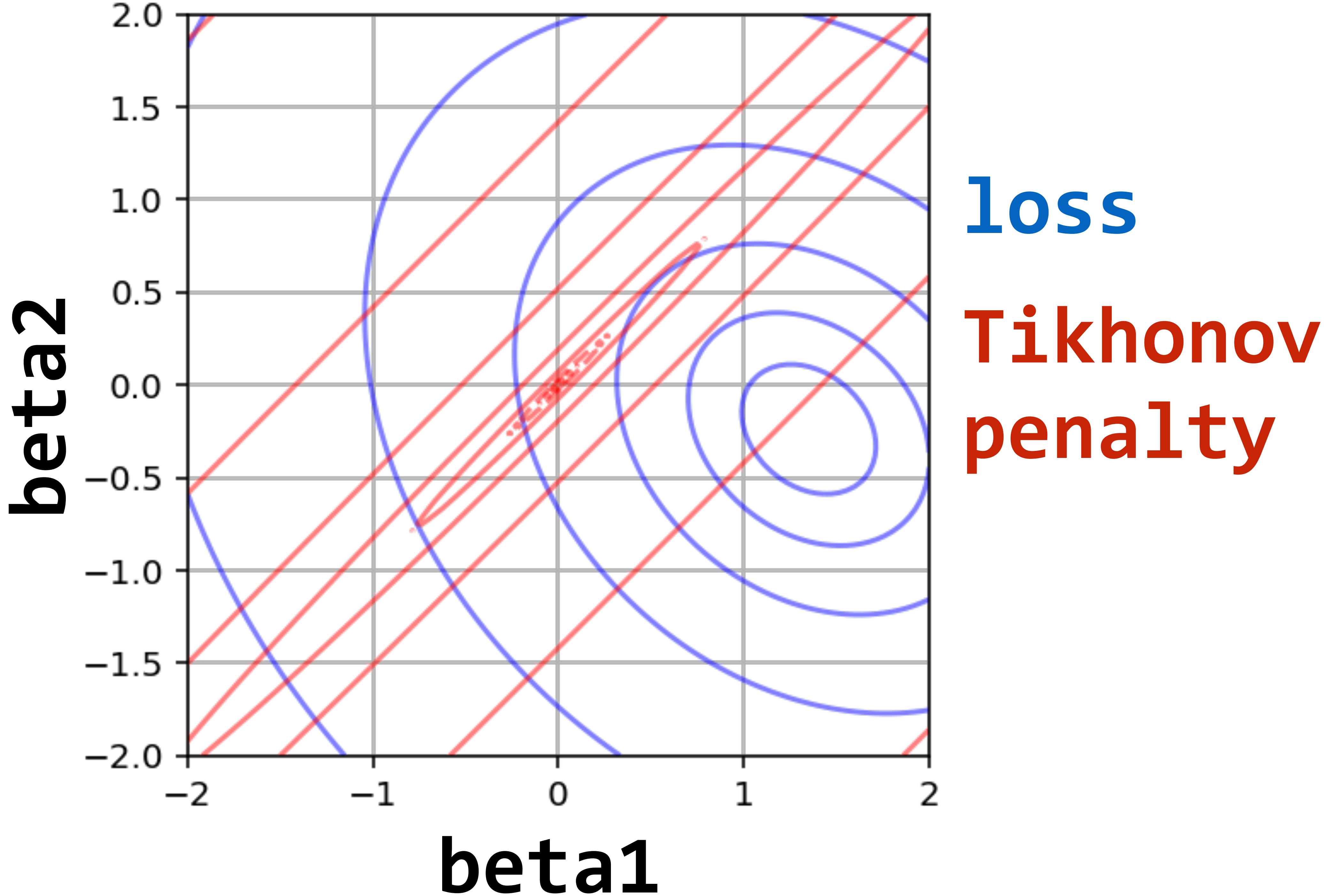


EXAMPLE

* Suppose we strongly suspect that **beta1** and **beta2** should be similar

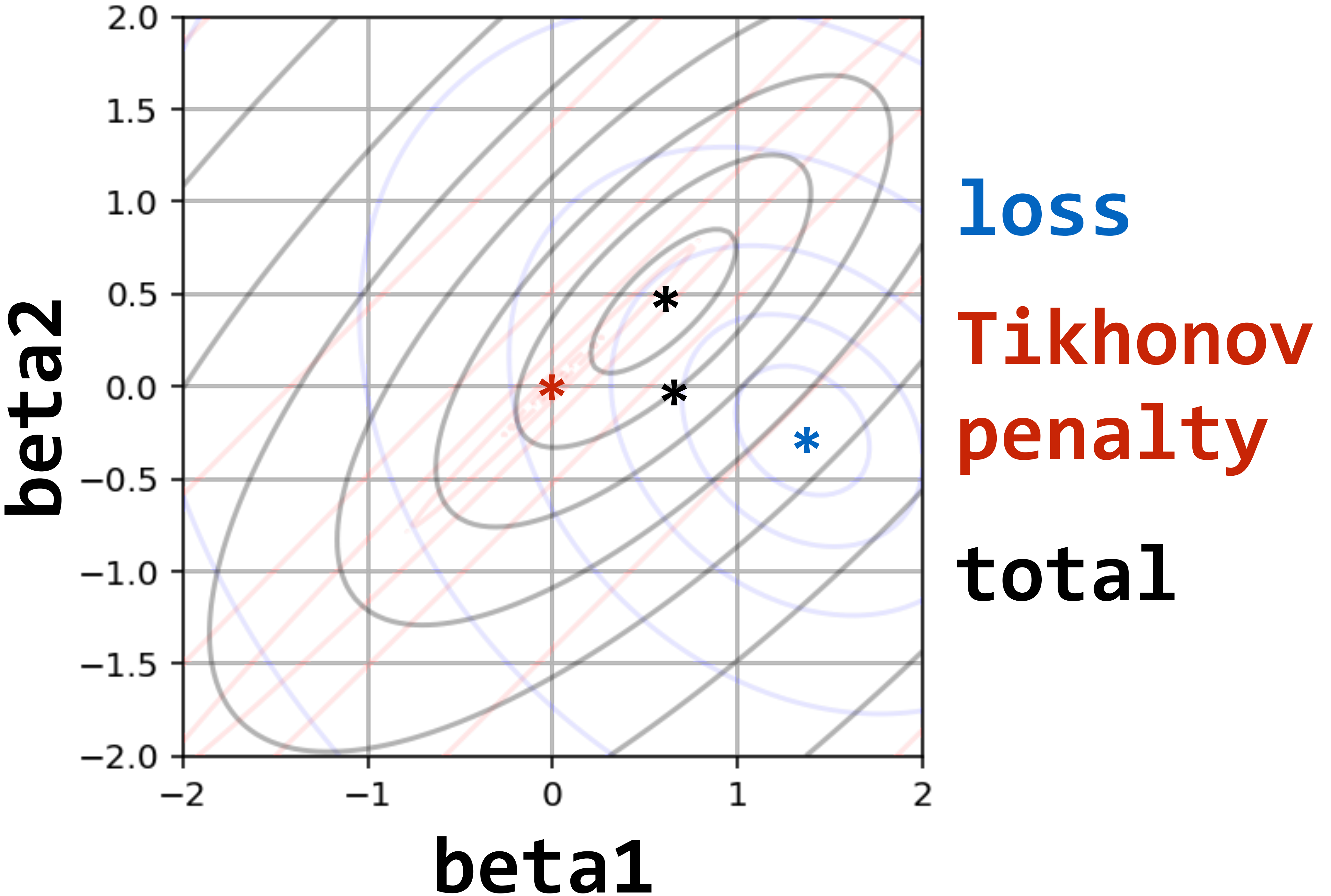
EXAMPLE

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



EXAMPLE

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



TIKHONOV REGRESSION

* any **TIKHONOV** problem can be converted into a **RIDGE** problem by a **LINEAR TRANSFORMATION**

* conversely, **ANY LINEAR TRANSFORMATION** of X followed by **RIDGE REGRESSION** is equivalent to some **TIKHONOV REGRESSION** problem

TIKHONOV REGRESSION

* WORD EMBEDDING MODELS

* think of stimulus matrix as **WORDS** over time projected onto WORD EMBEDDING

WORDS EMBEDDING

$$\underset{t\times n}{X} = \underset{t\times w}{W} \underset{w\times n}{E}$$

TIME WORDS EMBEDDING DIMENSIONS

\underset{t\times n}{X} = \underset{t \times w}{W} \underset{w\times n}{E}

TIKHONOV REGRESSION

* this is equivalent to TIKHONOV REGRESSION on the **WORDS** with a prior determined by the WORD EMBEDDING

$$\frac{1}{\sigma^2}\Sigma_\beta = (C^T C)^{-1} = E^T E$$

PRIOR COVARIANCE

INVERSE OF PENALTY INNER PRODUCT

EMBEDDING INNER PRODUCT

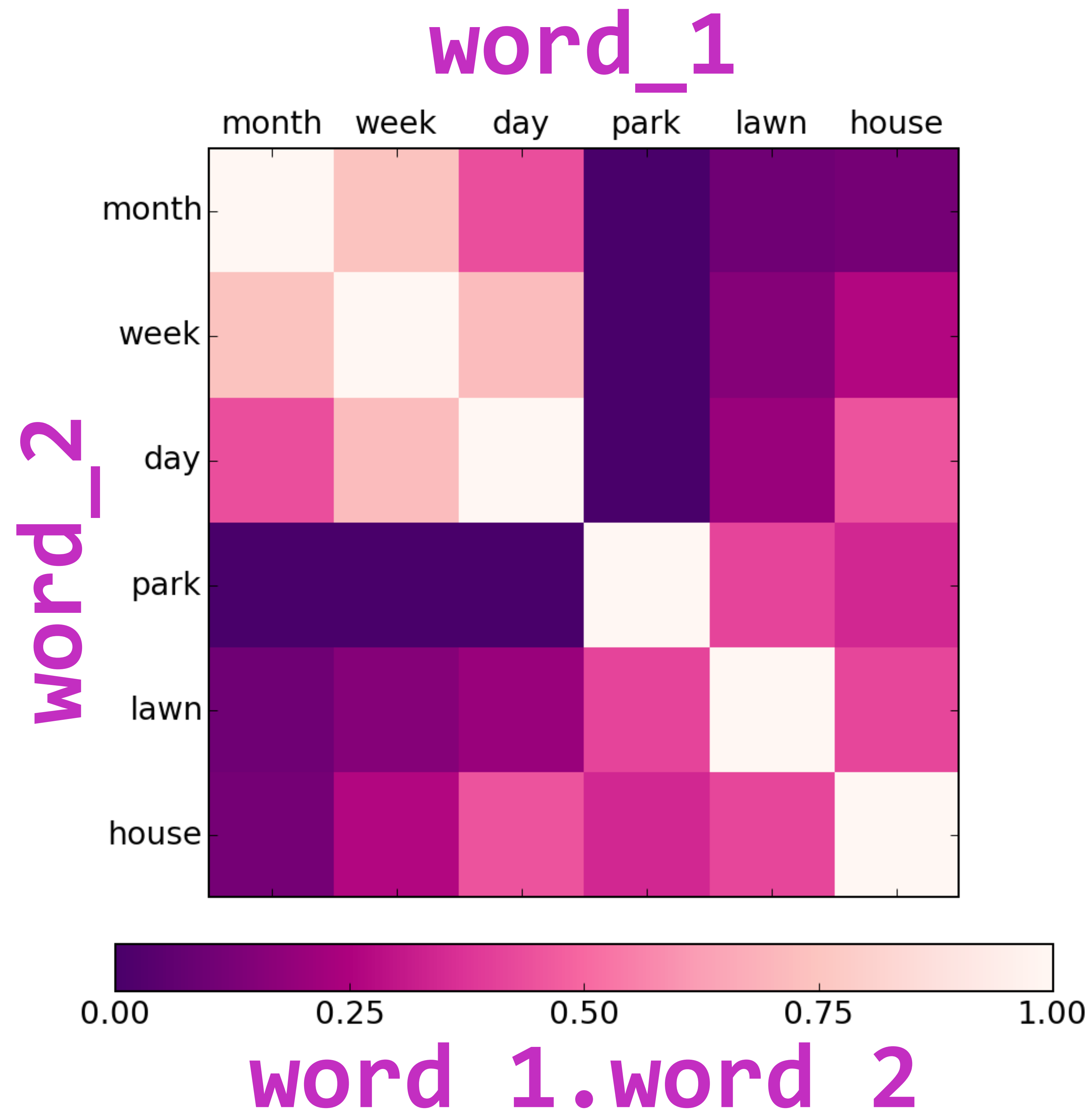
* **i.e. the prior covariance between two words' weights is equal to the dot product of their embedding vectors**

\frac{1}{ \sigma^2}
\Lambda^{-1} = E^

TIKHONOV REGRESSION

$$E^T E =$$

EMBEDDING INNER PRODUCT, ***english1000***



TIKHONOV REGRESSION

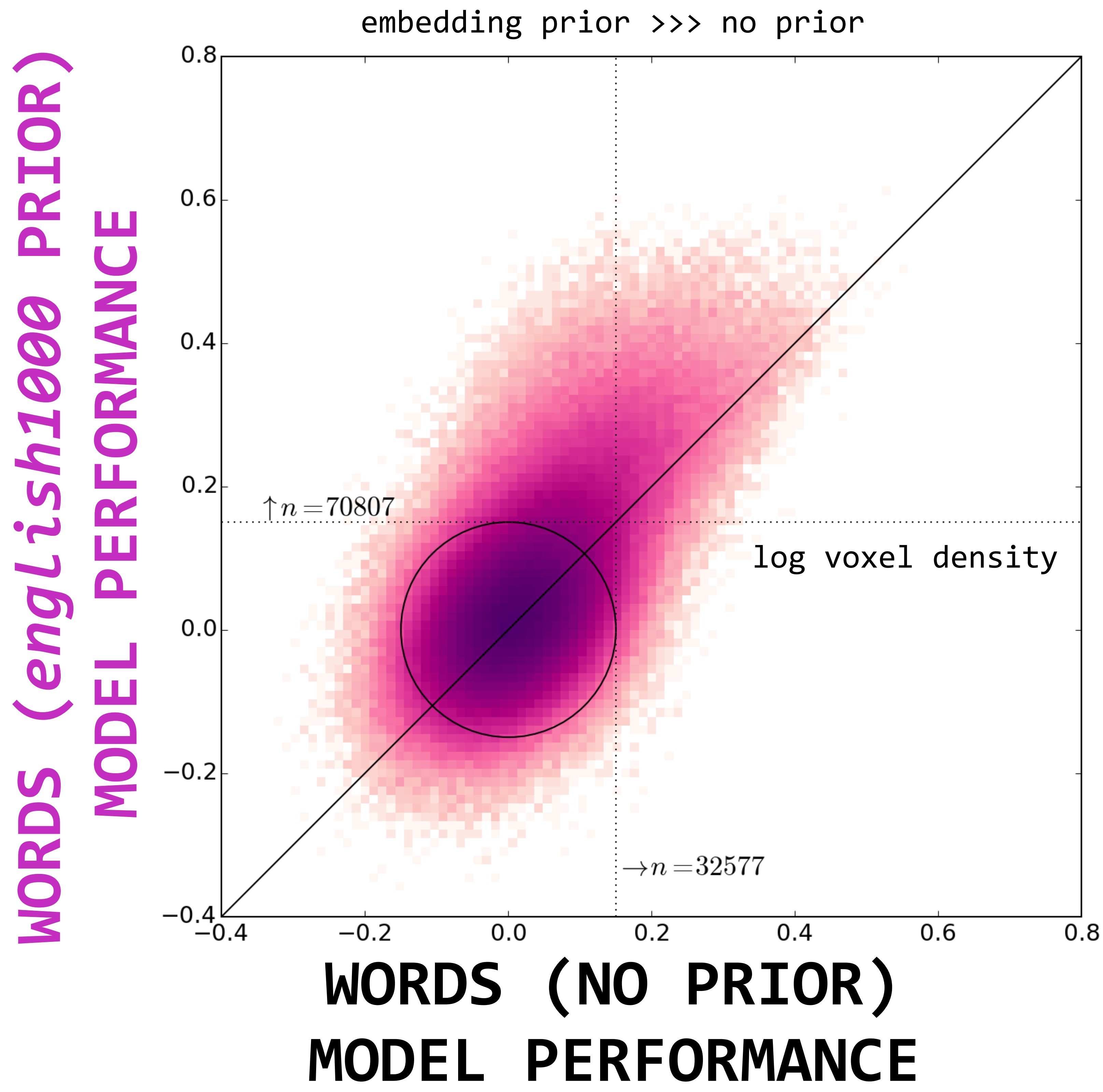
\underset{w \times v}{\
\underset{w \times n}{
\underset{n \times v}{\

* to get **WEIGHTS ON WORDS** we just project onto the **EMBEDDING**

$$\underset{w \times v}{\overset{\text{WEIGHTS IN WORD SPACE}}{\hat{\beta}_W}} = \underset{w \times n}{\overset{\text{EMBEDDING}}{E}} \; \underset{n \times v}{\overset{\text{WEIGHTS IN EMBEDDING SPACE}}{\hat{\beta}_X}}$$

* (this is equivalent to simulating responses to single words)

TIKHONOV REGRESSION



NEXT TIME

* Data quality!