rodluger@gmail.com

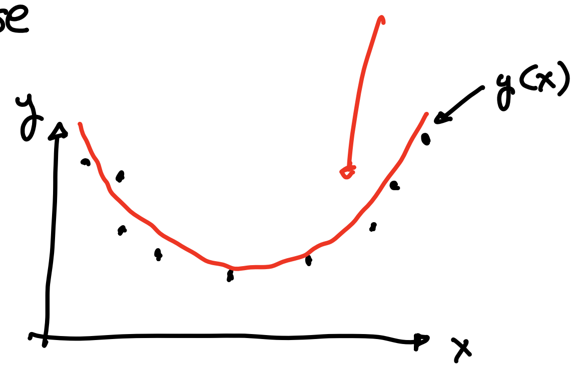
OLS

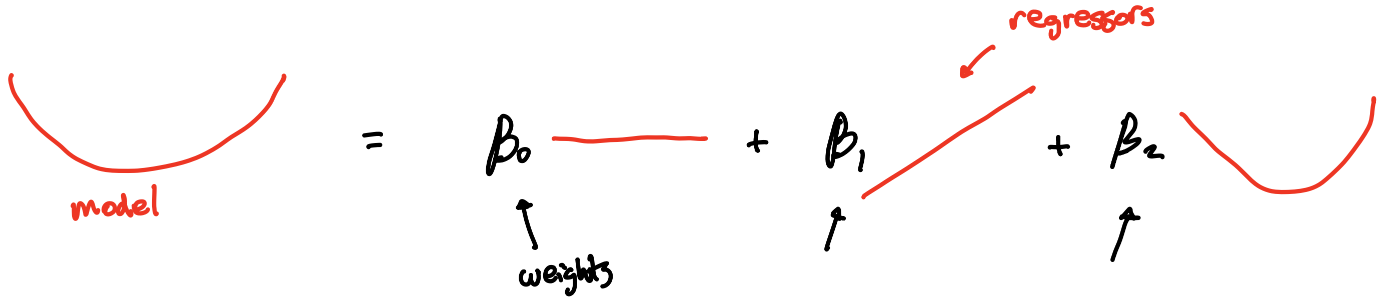
weights

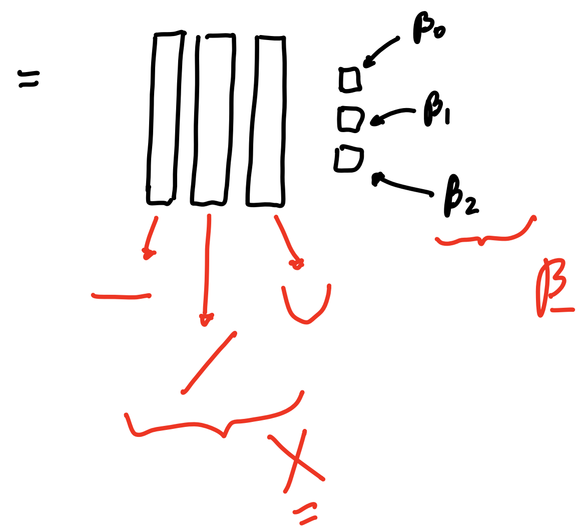
$$\underline{y} = \underline{\underline{X}}\,\underline{\beta} + \text{noise}$$

data

design matrix







$$\rightarrow \quad \underline{y} = \underline{\underline{X}}\,\underline{\beta} \quad \leftarrow \qquad \underline{\underline{X}}^T \underline{y} = \underline{\underline{X}}^T \underline{\underline{X}}\,\underline{\beta}$$

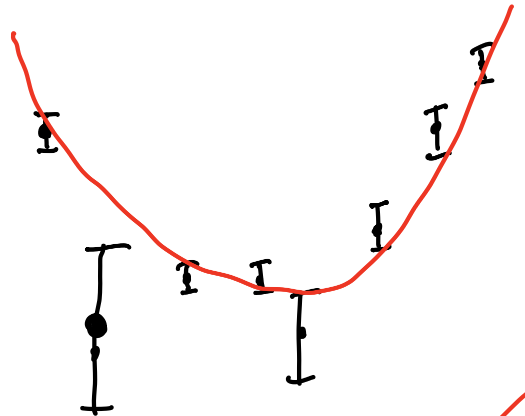
$$\cancel{\hat{\beta} \overset{?}{=} \underline{\underline{X}}^{-1}\underline{y}}$$

does not work

$$\boxed{\left(\underline{\underline{X}}^T \underline{\underline{X}}\right)^{-1} \underline{\underline{X}}^T \underline{y} = \underline{\hat{\beta}}_{OLS}}$$

best fit model $= \underline{\underline{X}} \cdot \underline{\hat{\beta}}_{OLS}$

GLS



not constant

$$y = X\beta + \text{noise}$$

$$\Sigma = \begin{bmatrix}\sigma_0^2 & & & \\ & \sigma_1^2 & & \\ & & \ddots & \\ & & & \sigma_{n-1}^2\end{bmatrix}$$

$$\to \quad \ln P = -\frac{1}{2}(y - X\beta)^T \Sigma^{-1} (y - X\beta) + \text{const}$$

$$\sim \chi^2 = \sum_i \left(\frac{\text{data}_i - \text{model}}{\text{uncert}}\right)^2$$

$$\to \quad \frac{d \ln P}{d\beta} = 0 \text{, solve for } \beta$$

(optional homework)

$$\boxed{\hat{\beta}_{GLS} = \underbrace{\boxed{(X^T \Sigma^{-1} X)^{-1}}}_{= \Sigma_{\beta_{GLS}}} X^T \Sigma^{-1} y}$$

$$\boxed{\beta \sim \mathcal{N}(\hat{\beta}, \Sigma_{\beta_{GLS}})}$$

POSTERIOR

DATA

POSTERIOR $\propto$ LIKELIHOOD $\times$ PRIOR

$$\ln P = \ln \mathcal{L} + \ln \text{Prior} + \text{const}$$

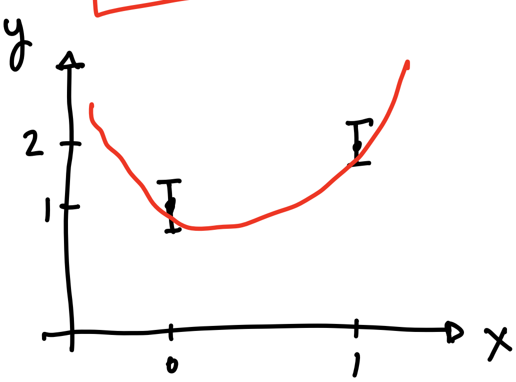
$$\ln P = \overbrace{-\tfrac{1}{2}(y - X\beta)^T \Sigma^{-1} (y - X\beta)}^{\ln \mathcal{L}} + \text{const}$$

GLS

prior on β is $U[-\infty, \infty]$

WHY BOTHER w/ A PRIOR?

B/c $\boxed{X^T \Sigma^{-1} X}$ is NOT ALWAYS INVERTIBLE!



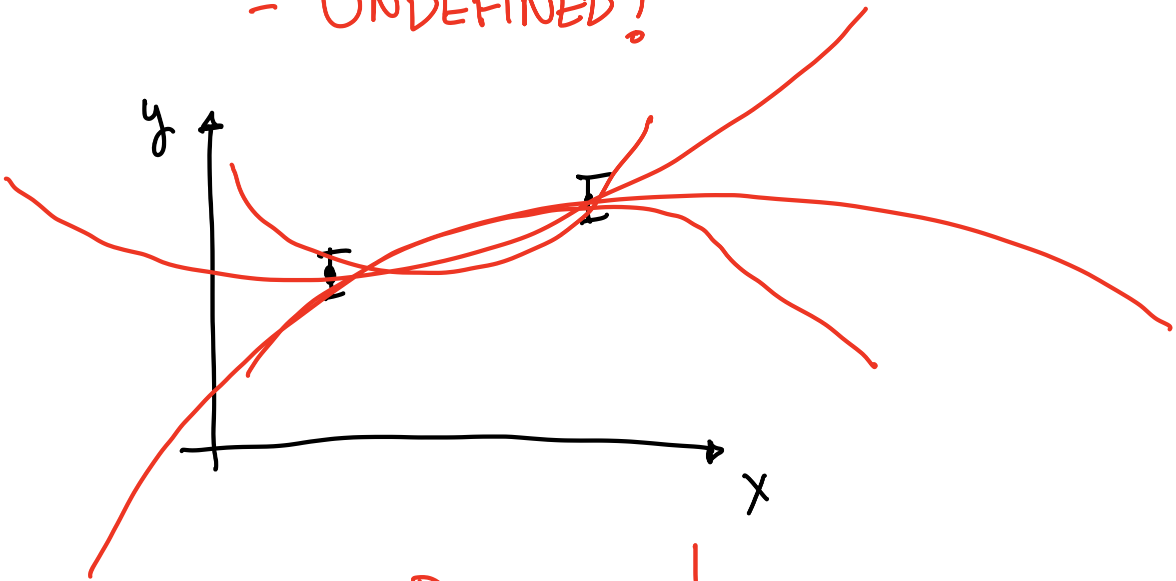
$$y = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad \Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$y = X\beta$$

$$X = \begin{array}{ccc} \text{const} & x & x^2 \\ \end{array}$$
$$X = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

$$X^T \Sigma^{-1} X = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

= UNDEFINED!



DEGENERACIES!

DATA IS NOT SUFFICIENTLY INFORMATIVE.

REGULARIZATION

(L2 NORM, RIDGE REGRESSION)

$$\ln P = -\frac{1}{2}(y - X\beta)^T \Sigma^{-1} (y - X\beta) + \underline{\underline{\text{prior on } \beta}} + \text{const}$$

$\underline{\underline{\text{Prior on } \beta}}$:

$$\beta \sim \mathcal{N}(\underbrace{\beta_0}_{\text{prior mean}}, \underbrace{C}_{\text{prior covariance}})$$

$$-\frac{1}{2}(\beta - \beta_0)^T C^{-1} (\beta - \beta_0) + \text{const}$$

$$= -\frac{1}{2}\beta^T C^{-1} \beta \qquad (\beta_0 = 0)$$

Scalar

$$= -\frac{\lambda}{2}\beta^T\beta \qquad (C = \lambda^{-1} I)$$

identity

$$= \boxed{-\frac{\lambda}{2}\sum_i \beta_i^2}$$

L2 PENALTY

GLS :

$$\hat{\beta}_{GLS} = \left(X^T\Sigma^{-1}X\right)^{-1}X^T\Sigma^{-1}y$$

L2 :

$$\hat{\beta}_{L2} = \left(\underline{X^T\Sigma^{-1}X} + \lambda I\right)^{-1}X^T\Sigma^{-1}y$$

stability

L2 REGULARIZATION $\Leftrightarrow$ GAUSSIAN PRIOR ON β

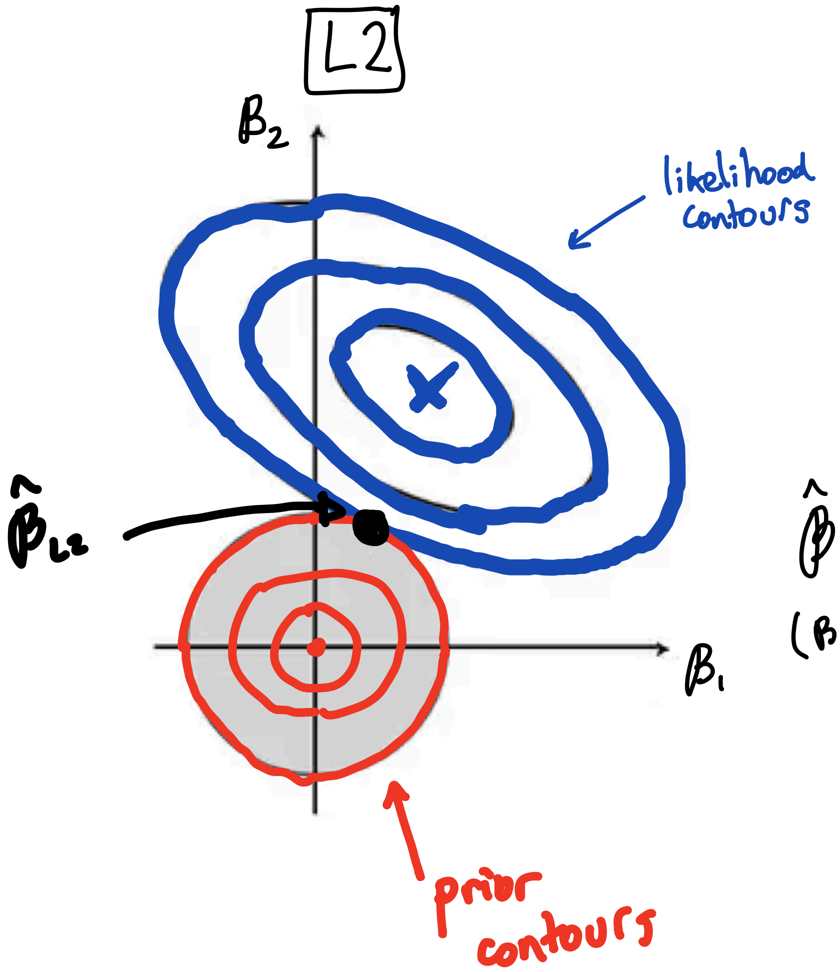
L1 : (LASSO)

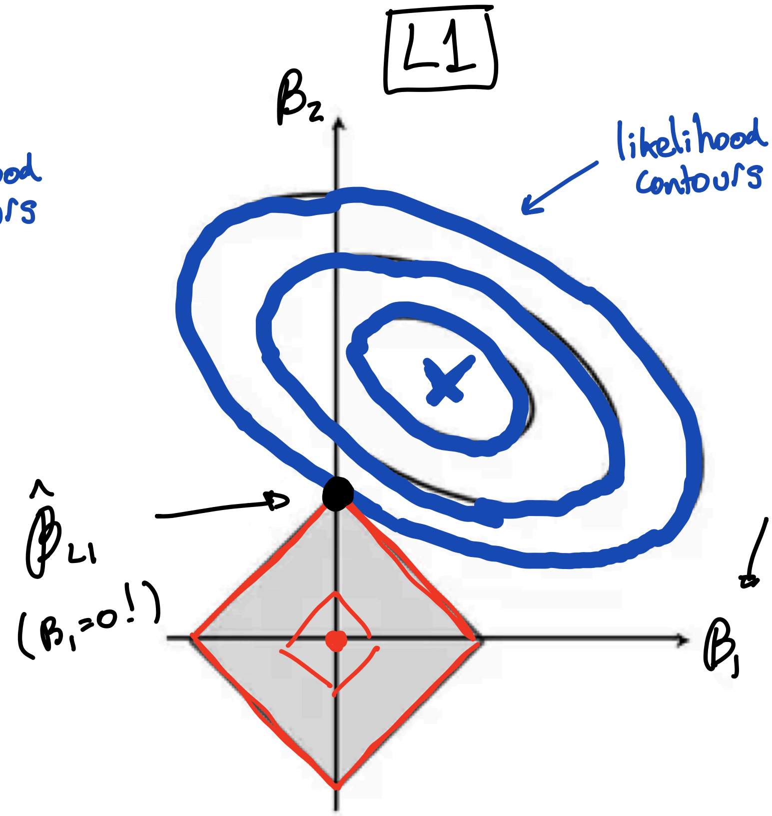
$$L2: \quad -\frac{\lambda}{2}\sum_i \beta_i^2$$

$$L1: \quad -\frac{\lambda}{2}\sum_i |\beta_i|$$

(NO CLOSED FORM SOLUTION)

BIASED!

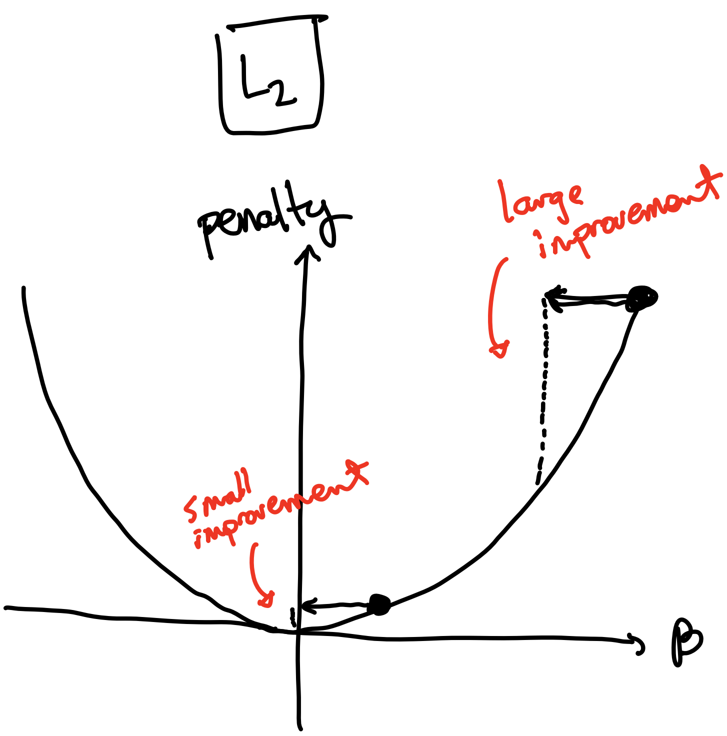


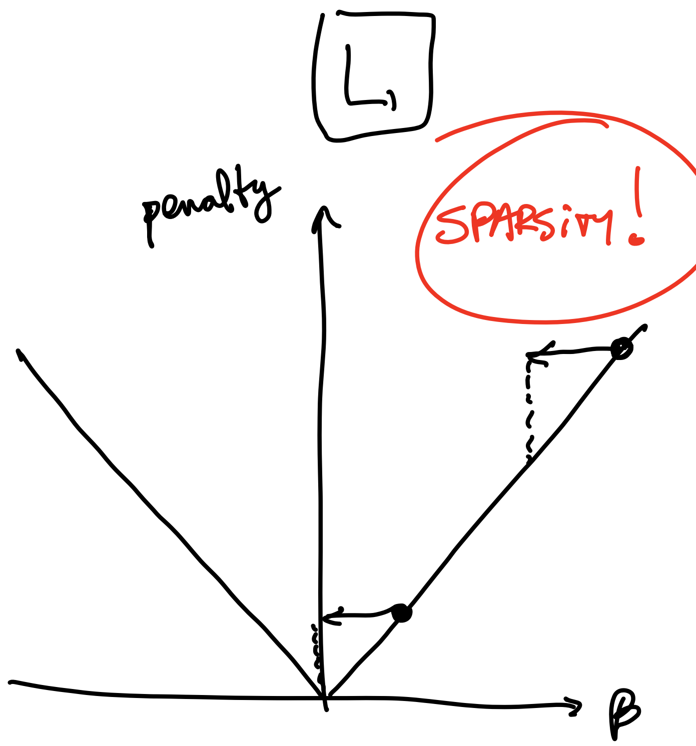


$$-\frac{\lambda}{2} \underbrace{\sum_i |\beta_i|}$$

$$\beta_1 + \beta_2 = \text{const}$$

$$\beta_2 = \text{const} - \beta_1$$





L2	L1
$\text{loss} \mathrel{+}= \lambda \sum_i \beta_i^2$	$\text{loss} \mathrel{+}= \lambda \sum_i \lvert \beta_i \rvert$
NOT SPARSE	SPARSE SOLUTION
UNIQUE SOLUTION	NOT NECESSARILY UNIQUE
NO	BUILT-IN FEATURE SELECTION
ISSUES W/ OUTLIERS	MORE ROBUST TO OUTLIERS
BAYESIAN INTERPRETATION	NO
CLOSED FORM SOL'N (ONLY WHEN PROB IS LINEAR)	NO