Here we introduce another method: **Ridge regression**

In the bayesian world, you would need to infer β parameter values, which you would then not go for the ML method (frequentist approach). Instead you would ask what is the most probable, likely β, that is most probable for seeing your data?

$$Z = \{(x_i, y_i) \quad 1 \le i \le n\}, \text{ Full set of observations}$$

$$= \{z_1, \dots, z_n \quad 1 \le i \le n\}$$

Bayes formula: $P(\beta|Z) = \dfrac{P(Z|\beta) \cdot P(\beta)}{P(Z)}$ (likelihood fxn, prior, evidence)

Fixed data Z.

Then you might go for: **MAXIMUM a POSTERIORI ESTIMATE (MAP)**

$$\hat{\beta}_{MAP} \in \arg\max_{\beta} P(\beta|Z) \qquad \text{posterior probability } P(\beta|Z).$$

$$= \arg\min_{\beta} \{ -\log P(Z|\beta) - \log P(\beta) \}$$ (likelihood) minimizing ~~cost~~ fxn. chosen model

* maximization of a likelihood fxn under conditional Gaussian noise distribution for a linear model is the same to minimize a sum-of-squares error fxn.

REGULARIZATION

Before you look at the data, know that we assumed β to be distributed by nature according to gaussian law.

How to choose $P(\beta)$? prior

β is distributed by nature, $P(\beta) = \prod_{i=1}^{d} \dfrac{\exp\left(\frac{-\beta^2}{2\sigma_\beta^2}\right)}{\sqrt{2\pi\sigma_\beta^2}}$

Take the log of $P(\beta)$.

$$-\log P(\beta) = \sum_{j=1}^{d} \frac{\beta^2}{2\sigma_\beta^2} + \text{constant}$$

d = dim
n = examples

$$= \arg\min_{\beta} \left\{ \sum_{i=1}^{n} (Y_i - X_i^T \beta)^2 + \lambda \sum_{j=1}^{d} \beta^2 \right\}$$

// cost fxn for ridge regression

$(Y_i - X_i^T\beta)$: prediction; whole sum: error term; $\lambda \sum \beta^2$: penalty regularizer, scalar to penalize max-likelihood fxn

used TO control model complexity, its an isotropic type of penalty term for choosing the number of β's you many want to restrict.

cost fxn depends on β.

$$RSS(\beta) = (Y - \vec{X}\beta)^T (Y - X\beta) + \lambda \|\beta\|^2$$

data term; penalty term

where λ = lagrange parameter

$\lambda = \dfrac{1}{2\sigma_\beta^2}$, validated by data and controls solution to fitting parameters of the model.