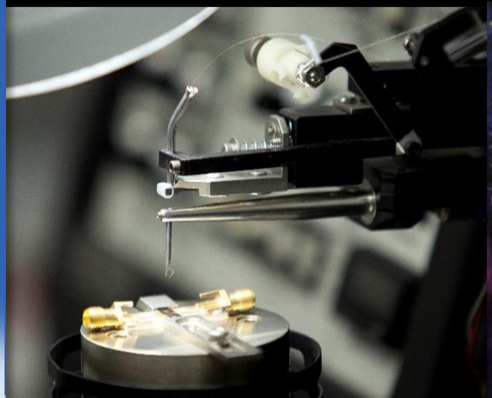
Centro Brasileiro de Pesquisas Físicas

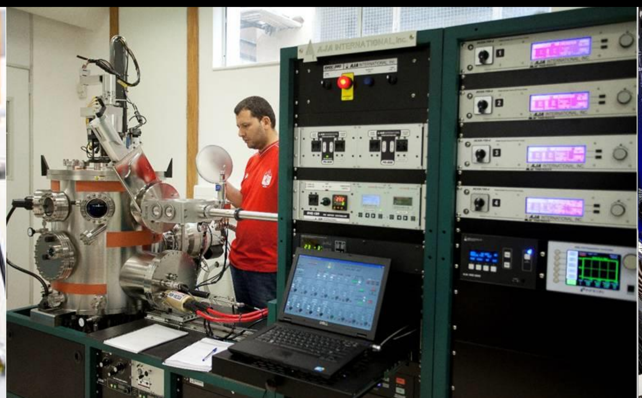
Métodos para Análise de grande volume de dados e Astroinformática

Clécio Roque De Bom – debom@cbpf.br

CBPF

clearnightsrthebest.com





Quick Guide to MLE

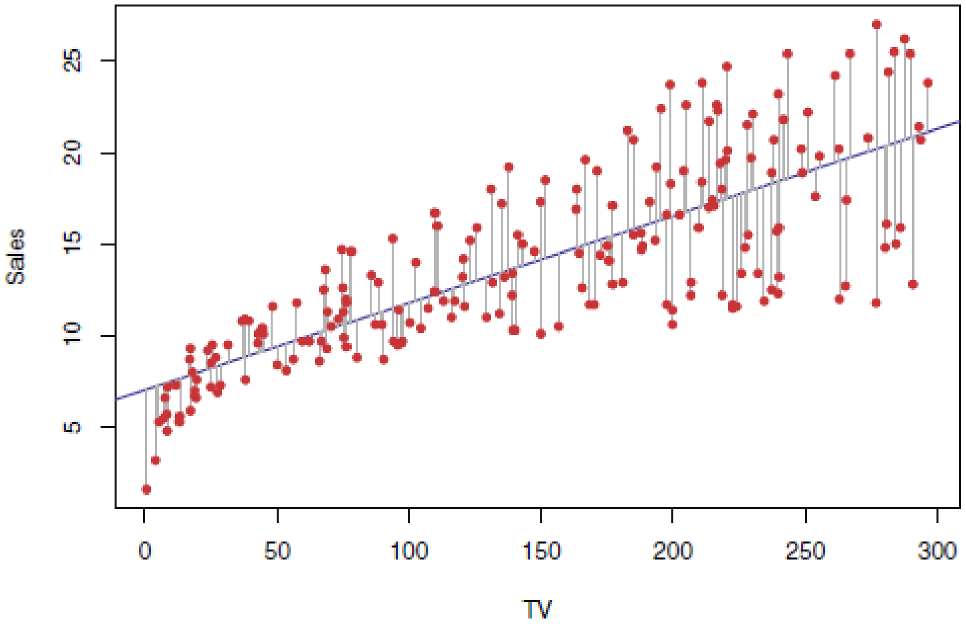
What is the distribution that represents your stochastic variable?

Is it Gaussian? Poisson? Bernulli?

What is the inner model that produces those variables?

Is it linear? Polynomial? Is it a exponential?

If your variable is idd, then get the logP and find the most probable parameters by using your preferred minimization method.



$$Y = f^*(X) + \epsilon = X\beta^* + \epsilon$$

$$\epsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{I}) \qquad Y \sim \mathcal{N}(X\beta^*, \sigma^2 \mathbf{I})$$

$$\widehat{\beta}_{\text{MLE}} = \arg\max_{\beta} \log p(\{(X_i, Y_i)\}_{i=1}^n | \beta, \sigma^2)$$

$$= \arg\min_{\beta} \sum_{i=1}^{n} (X_i\beta - Y_i)^2 = \widehat{\beta}$$

Logistic Regression

$$P(y = 1 \mid X) = p$$

$$\text{odds} = \frac{\text{probability of something happening}}{\text{probability of something not happening}} = \frac{p}{1-p}$$

Logistic Regression

$$\ln\left(\frac{P}{1-P}\right) = \theta_1 + \theta_2 x + \mathrm{e}$$

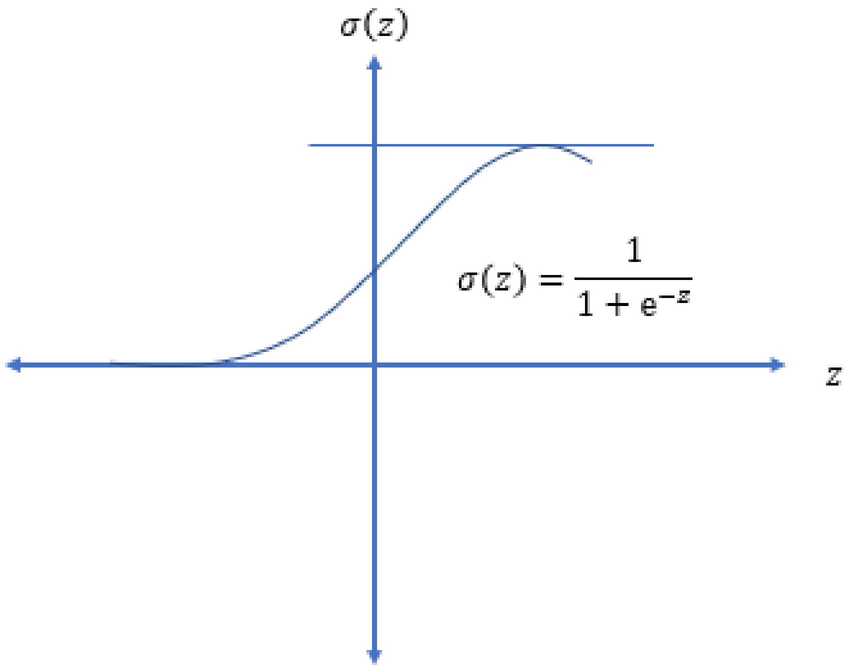
$$\frac{P}{1-P} = \mathrm{e}^{\theta_1+\theta_2 x+\mathrm{e}}$$

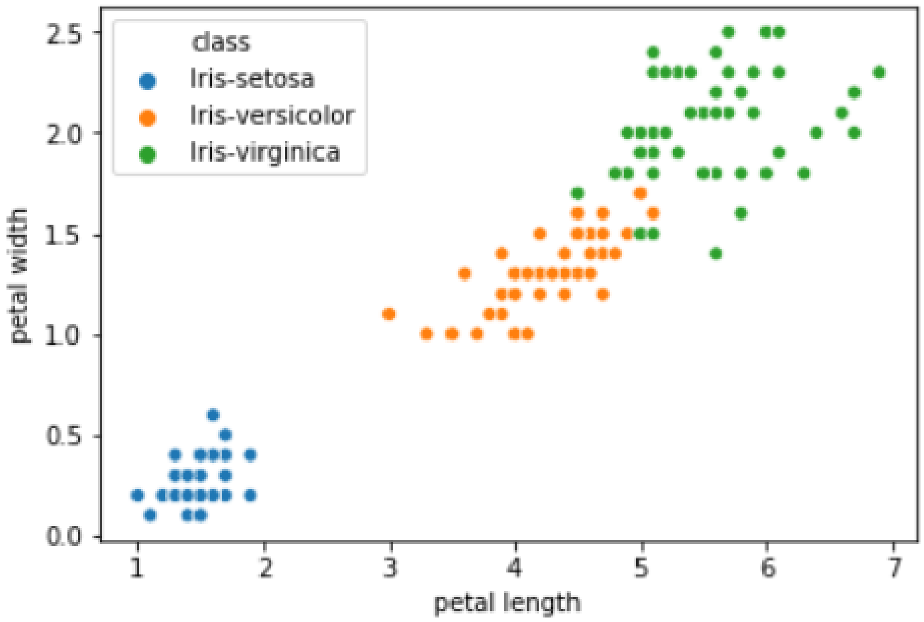
$$P = \frac{1}{1+\mathrm{e}^{-}}(\theta_1 + \theta_2 x)$$

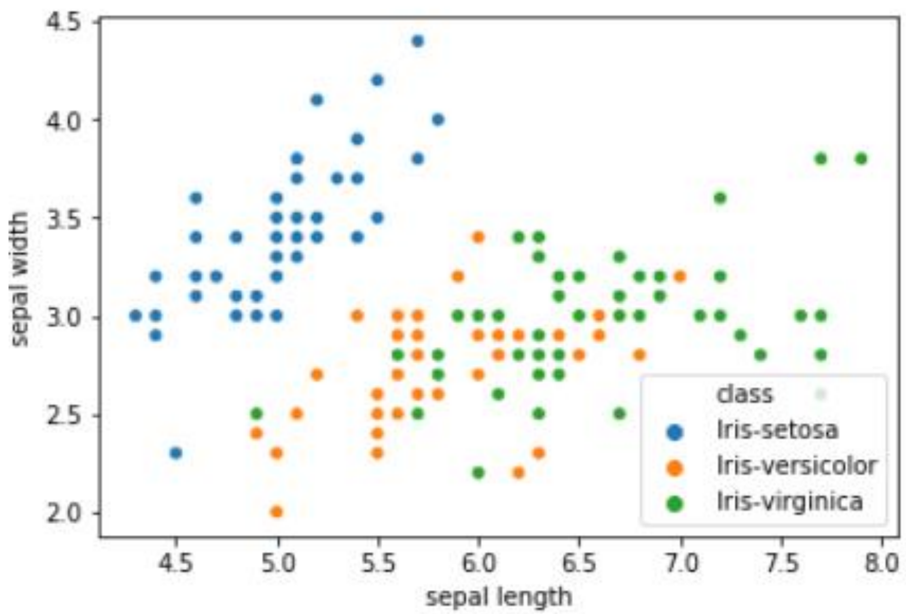
$$\sigma(z) = \frac{1}{1+\mathrm{e}^{-z}} \quad \text{Where} \quad z = \theta^T x$$

$$\theta^T \mathbf{x} = \sum_{i=1}^{m} \theta_i x_i = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_m x_m$$

Logistic Regression





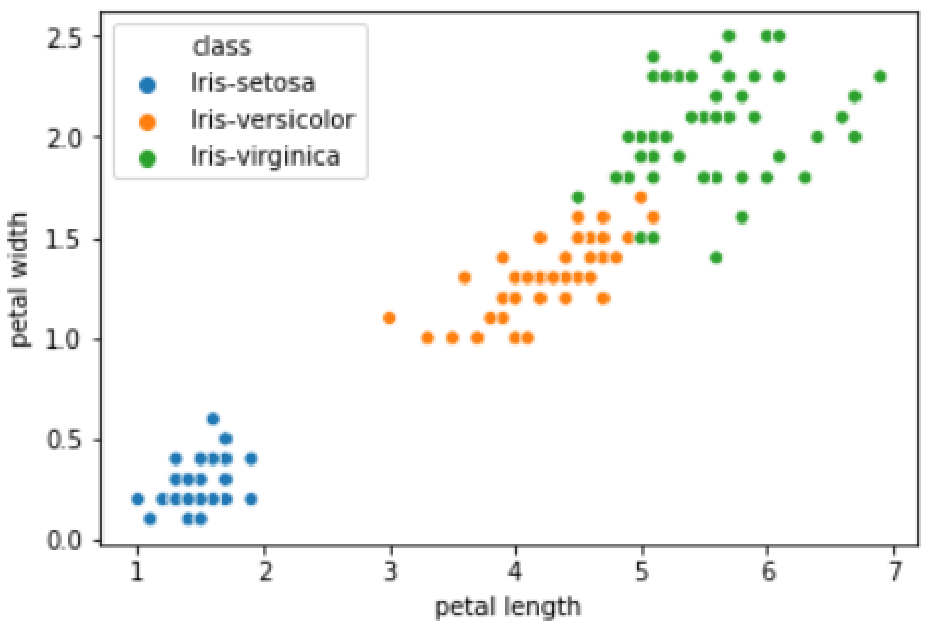


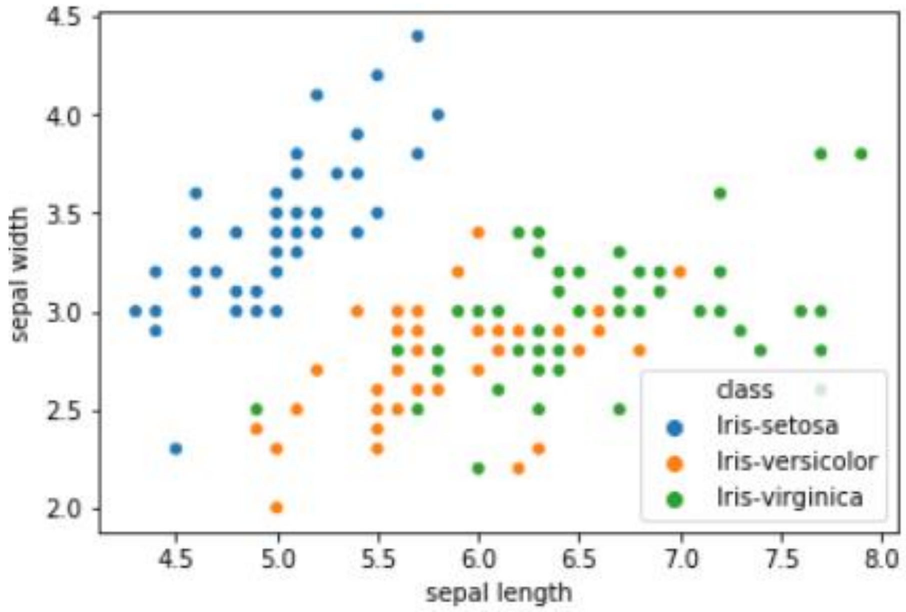
Logistic Regression

$$P(Y = 1 \mid \mathbf{X} = \mathbf{x}) = \sigma(\theta^T \mathbf{x})$$
$$P(Y = 0 \mid \mathbf{X} = \mathbf{x}) = 1 - \sigma(\theta^T \mathbf{x})$$

$$P(Y = y|X = \mathbf{x}) = \sigma(\theta^T \mathbf{x})^y \cdot \left[1 - \sigma(\theta^T \mathbf{x})\right]^{(1-y)}$$

$$LL(\theta) = \sum_{i=1}^{n} y^{(i)} \log \sigma(\theta^T \mathbf{x}^{(i)}) + (1 - y^{(i)}) \log[1 - \sigma(\theta^T \mathbf{x}^{(i)})]$$





Quick Guide to MLE

What is the distribution that represents your stochastic variable?

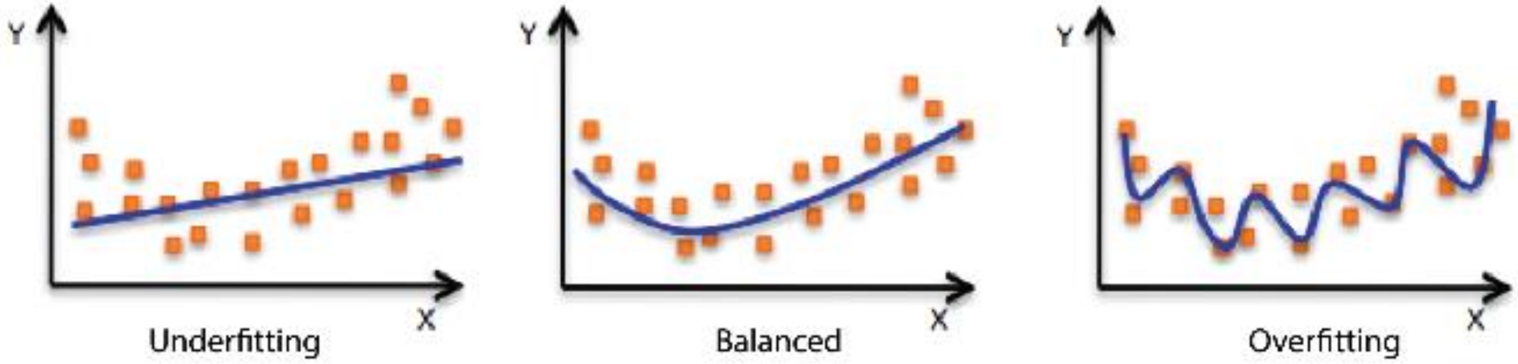
Is it Gaussian? Poisson? Bernulli?

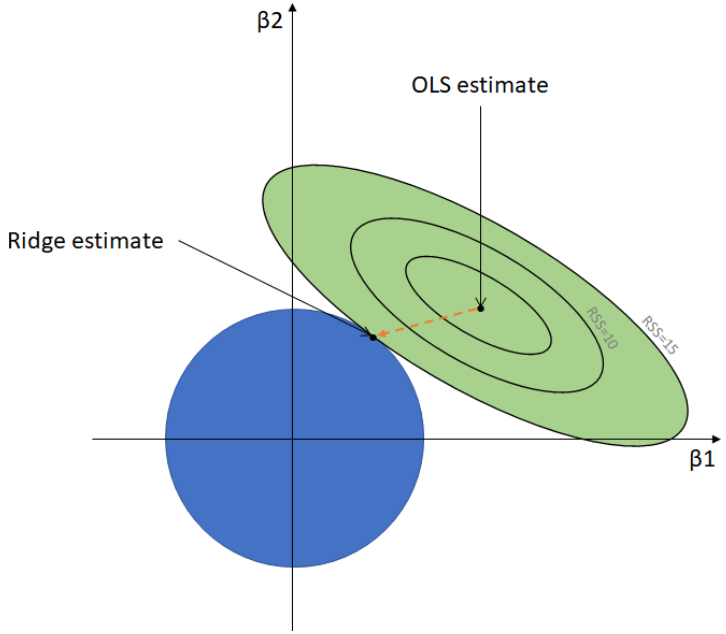
What is the inner model that produces those variables?

Is it linear? Polynomial? Is it a exponential?

If your variable is idd, then get the logP and find the most probable parameters by using your preferred minimization method.

Robustness – Example Ridge regression





$$\underset{\beta\in\mathbb{R}}{argmin}\sum[y_i - \hat{y}_i] = \underset{\beta\in\mathbb{R}}{argmin}\sum[y_i - (\beta_0 + \beta_1 x_1 + \beta x_2 + \cdots + \beta x_p)]^2$$

$$\beta_0^2 + \beta_1^2 + \cdots + \beta_p^2 \leq C^2$$

$$\|B\|_2 = \sqrt{\beta_0^2 + \beta_1^2 + \cdots + \beta_p^2}$$

$$\hat{\beta}^{ridge} = \underset{\beta\in\mathbb{R}}{argmin}\|y - XB\|_2^2 + \lambda\|B\|_2^2$$

Adapted from https://towardsdatascience.com/ridge-regression-for-better-usage-2f19b3a202db

Ministério da Ciência, Tecnologia e Inovações

Centro Brasileiro de Pesquisas Físicas

Métodos para Análise de grande volume de dados e Astroinformática

Clécio Roque De Bom – debom@cbpf.br

clearnightsrthebest.com

CBPF