

Introduction to the 2nd Part of ASM: **Beyond LM and GLM**

Pedro Delicado

Departament d'Estadística i Investigació Operativa
Universitat Politècnica de Catalunya

In the first part of Advanced Statistical Models

- Multiple Linear Regression Model (LM):

- $Y = \mathbf{x}^T \boldsymbol{\beta} + \varepsilon$, $Y \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^p$.
- Data: (Y_i, \mathbf{x}_i) , $i = 1, \dots, n$.
- Estimation: Ordinary Least Squares (OLS),

$$\hat{\boldsymbol{\beta}}_{\text{OLS}} = \arg \min_{\boldsymbol{\beta}} \sum_{i=1}^n \left(Y_i - \mathbf{x}_i^T \boldsymbol{\beta} \right)^2.$$

- Generalized Linear Model (LM):

- $(Y|\mathbf{X} = \mathbf{x}) \sim f(y; \mu(\mathbf{x}))$, $Y \in S \subseteq \mathbb{R}, \mathbf{x} \in \mathbb{R}^p$,
 $\mu(\mathbf{x}) = \mathbb{E}(Y|\mathbf{X} = \mathbf{x})$, $g(\mu(\mathbf{x})) = \mathbf{x}^T \boldsymbol{\beta}$, $\mu(\mathbf{x}) = g^{-1}(\mathbf{x}^T \boldsymbol{\beta})$.
- Data: (Y_i, \mathbf{x}_i) , $i = 1, \dots, n$, independent.
- Estimation: Maximum Likelihood,

$$\hat{\boldsymbol{\beta}}_{\text{ML}} = \arg \max_{\boldsymbol{\beta}} \sum_{i=1}^n \log \left(f(Y_i; g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta})) \right).$$

In the second part of Advanced Statistical Models

- First, we explore **penalized** estimation criteria:
 - Ridge regression and Lasso for LM and for GLM.
- Second, we remove the assumptions of linearity. In fact we are not considering parametric forms for $m(\mathbf{x}) = \mathbb{E}(Y|\mathbf{X} = \mathbf{x})$ (**nonparametric regression**):
 - Local linear regression, local likelihood estimators.
 - Spline smoothing.
 - Generalized Additive Models (GAM).

Contents of 2nd part of the course

Regularized estimation of LM and GLM. 1. Ridge regression. 2. The Lasso estimation. 3. Lasso estimation in the GLM.

Nonparametric regression model. 1. Local polynomial regression. 2. Kernel functions. 3. Theoretical properties. The bias-variance trade off. 4. Bandwidth choice.

Generalized nonparametric regression model. 1. Nonparametric regression with binary response. 2. Generalized nonparametric regression model. 3. Estimation by maximum local likelihood.

Spline smoothing. 1. Penalized least squares nonparametric regression. 2. Splines, cubic splines and interpolation. 3. Smoothing splines. 4. B-splines and P-splines. 5. Spline regression. 6. Fitting generalized nonparametric regression models with splines.

Generalized additive models and Semiparametric models. 1. Multiple nonparametric regression. The curse of dimensionality. 2. Additive models. 3. Generalized additive models. 4. Semiparametric models.