ORF525 - Class Notes

Bachir EL KHADIR

February 11, 2016

Definition 1 (Oridnary Lease Squares Regression)**.** $f_i = \{f(x) = \beta^T X\}$ $\hat{\beta}^{OLS} = \arg\min_\beta ||Y - X\beta||_2^2$ $F(\beta) = Y^T Y + \beta^T X^T X \beta - 2\beta^T X^T Y$ $\frac{\partial F(\beta)}{\partial \beta} = 2X^T X\beta - 2X^T Y = 0 \implies \hat{\beta} = (X^T X)^{-1} X^T Y$

Definition 2 (Model-based Interpretation of OLS)**.** *Statistical Model* $Y = \beta^T X + \varepsilon, \varepsilon \sim \mathcal{N}(0,1)$ *Joint-Loglikelihood*

$$l_n(\beta, \sigma^2) = f \sum_{i=1}^n \log p_{\beta,\sigma^2}(Y_i, X_i) = \sum_{i=1}^n \log p_{\beta,\sigma^2}(Y_i|X_i) + \underbrace{\sum_{i=1}^n \log p(X_i)}_{\textit{does not depend on } \beta \textit{ or } \sigma^2}$$

$\implies$

$$\begin{aligned}\arg\max_{\beta,\sigma^2} l_n(\beta,\sigma^2) &= \arg\max_{\beta,\sigma^2} \underbrace{\sum_{i=1}^n \log p_{\beta,\sigma^2}(Y_i|X_i)}_{\textit{Conditional log-likelihood}} \\ &= \arg\max_{\beta,\sigma^2} \frac{1}{2\sigma^2} \sum (Y_i - \beta^T X_i)^2 + n\log(\frac{1}{\sqrt{2\pi\sigma^2}})\end{aligned}$$

$\implies \hat{\beta}^{MLE} = \arg\min \sum (Y_i - \beta^T X_i)^2 = \hat{\beta}^{OLS}$

1 Linear Regression with Basis Expansion

From linear to non linear

- Input vairables can be transofrmation of original feautres: Handraft features, Box-Cox tranformation (find the best transmformation)
- Input can have interactions, eg $X_1 X_2 \ldots$
- Inputs can have basis expansions. Instead of $f(x) = \beta^T x$ we can have $f(x) = \sum_j \beta_j \underbrace{h_j}_{\text{Adaptative learning}} (x)$.

Definition 3 (Categorical Variable)**.** *A variable that can take on only one of a limited values.* ***Dummy coding***

2 High Dimensional Regression

Definition 4 (High Dimensional Regression)**.** *Data when dimension d is bigger than the sample size n.*

$$Y = \begin{pmatrix} Y_1 \\ \cdots \\ Y_n \end{pmatrix}$$

$$X = \begin{pmatrix} X_{11} & \dots & X_{1n} \\ & \cdots & \\ X_{n1} & \dots & X_{nn} \end{pmatrix}$$

Question: $\hat{\beta}^{OLS} = (\underbrace{X^T X}_{\text{not invertible}})^{-1} X^T Y$, what should we do?

- Ridge Estimation $\hat{\beta}^{\lambda} = (\underbrace{X^T X + \lambda I}_{\text{Tuning Parameters}})^{-1} X^T Y \iff \hat{\beta}^{\lambda} = \arg\min_{\beta \in \mathbb{R}^d} ||Y - X\beta||_2^2 + \lambda ||\beta||_2^2 \iff \hat{\beta}^{\lambda} = \arg\min_{||\beta||_2^2 < t} ||Y - X\beta||_2^2$
-