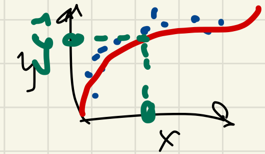
FoDA L14

Regression Misc.

Oct 6, 2022

Polynomial Regression



$M_\alpha(x) = \hat{y}$

Input $(x, y) \in \mathbb{R} \times \mathbb{R}$ $\quad \{(x_i, y_i)\}_{i=1}^{n}$ $\quad x_i \in \mathbb{R}$, $y_i \in \mathbb{R}$

$$x_i \rightarrow v_i = (1, x_i, x_i^2, \ldots, x_i^p) \in \mathbb{R}^{p+1}$$

$$\tilde{X}_p = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \in \mathbb{R}^{n \times (p+1)}$$

$y \in \mathbb{R}^n$

$$\alpha^* = (\tilde{X}_p^T \tilde{X}_p)^{-1} \tilde{X}_p^T y$$

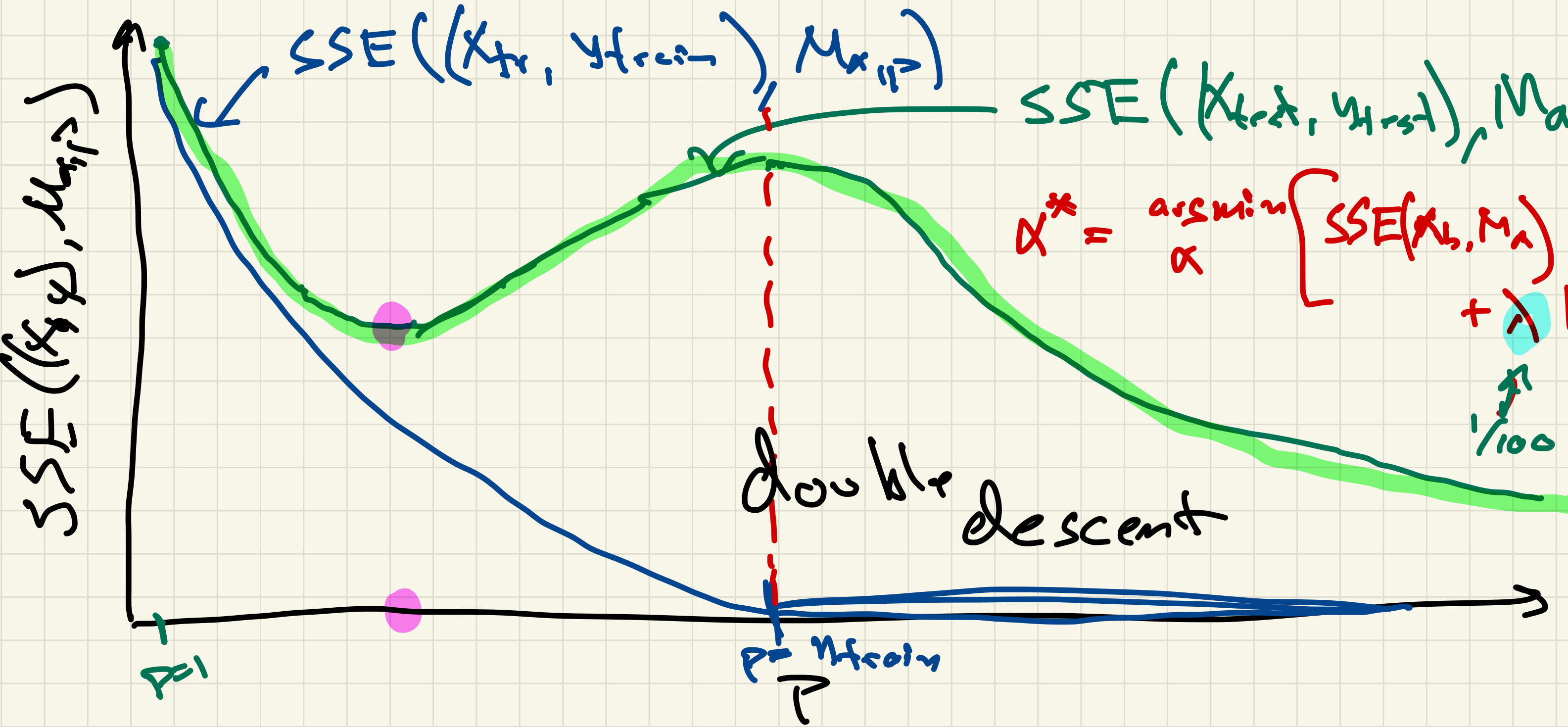
$$\alpha^* = (\alpha_0, \alpha_1 \ldots \alpha_p)$$

$$M_{\alpha^*, p}(x) = \langle \alpha^*, v \rangle$$

What is best value p?

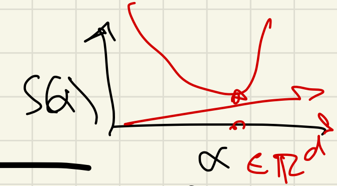
Choose best p for various p

$$(X, y) \xrightarrow{\text{random split}} \begin{cases} (X_{train}, y_{train}) \rightarrow \alpha^*_p \\ (X_{test}, y_{test}) \rightarrow \text{evaluate } SSE(test, \alpha) \end{cases}$$



$$\alpha^* = \underset{\alpha}{\operatorname{argmin}} \left[SSE(X_{tr}, M_\alpha) + \lambda \|\alpha\|^2 \right]$$

Why $\alpha^* = (X^TX)^{-1}X^Ty$?



$$S(\alpha) = SSE((x,y), M_\alpha) = \sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} \left(y_i - \langle \alpha, x_i \rangle\right)^2$$

fixed

$$r_i = \Big(y_i - \underbrace{\langle \alpha, x_i \rangle}_{\sum_{j=1}^{d} \alpha_j x_{ij}}\Big)$$

$$\frac{\partial r_i}{\partial \alpha_j} = -x_{ij}$$

$$0 = \frac{S(\alpha)}{\partial \alpha_j} = 2 \sum_{i=1}^{n} r_i \frac{\partial r_i}{\partial \alpha_j} = 2 \sum_{i=1}^{n} r_i (-x_{ij}) = \sum_{i=1}^{n} \left(y_i - \langle \alpha, x_i \rangle\right)(-x_{ij})$$

normal equations

$$\sum_{i=1}^{n} x_{ij} \langle x_i, \alpha \rangle = \sum_{i=1}^{n} x_{ij} y_i \quad \text{for all } j = 1 \dots d$$

$$(X^TX)^{-1}(X^TX)\alpha = X^Ty \Rightarrow \alpha = (X^TX)^{-1}X^Ty$$