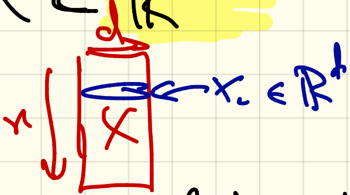
FoDA : Linear Regression

L14 : Recap

Supervised Learning Problem

Input (X, y) labels $\quad X \in \mathbb{R}^{\boxed{n \times d}} \quad y \in \mathbb{R}^n$



Goal: Build linear model to predict y. (actually on new data that looks like (X, y)).

estimate

$x \in \mathbb{R}^d \quad x = (x_1, x_2, \dots x_d)$

$$\hat{y} = M_\alpha(x) = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_d x_d$$

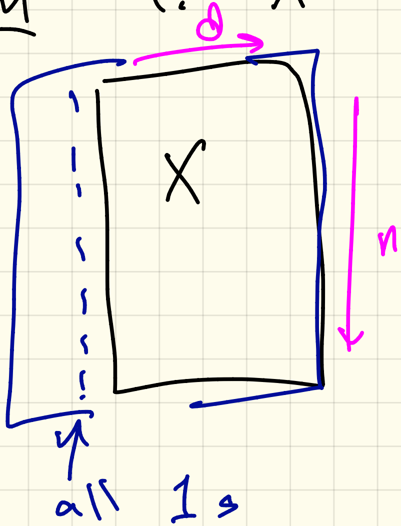
Goal: Find $\alpha = (\alpha_0, \dots \alpha_d)$ to minimize

$$SSE((X, y), \alpha) = \sum_{i=1}^{n} (M_\alpha(x_i) - y_i)^2$$

Soln 1. $X \to \tilde{X} \in \mathbb{R}^{n \times (d+1)}$

$$\tilde{x}_i = (1, x_i) \in \mathbb{R}^{d+1}$$

$\tilde{X} =$



Goal: $\alpha^* = \underset{\alpha \in \mathbb{R}^{d+1}}{\arg\min} \|\tilde{X}\alpha - y\|^2$

Soln #2: $\alpha = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T y$

often be written:

Goal: $\underset{\alpha}{\arg\min} \|X\alpha - y\|^2$

Soln: $\alpha^* = (X^T X)^{-1} X^T y$

Polynomial Regression

Input (x, y) $\quad x \in \mathbb{R}^n \quad y \in \mathbb{R}^n$

power

$x \in \mathbb{R}^1$

$$\hat{y} = M_\alpha^{(p)}(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \dots + \alpha_p x^p$$

$$= \sum_{j=0}^{p} \alpha_j x^j$$

Soln

1. $\tilde{X}_p \in \mathbb{R}^{n \times (p+1)}$

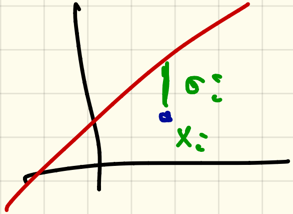
$$\tilde{X}_p = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^p \\ 1 & x_2 & x_2^2 & \dots & x_2^p \\ & & & & \\ 1 & x_n & x_n^2 & \dots & x_n^p \end{bmatrix}$$

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

Cross-Validation

(specific for LinReg)

1. residuals : error $\displaystyle\sum_{i=1}^{n} r_i^2$



$r_i = M_\alpha(x_i) - y_i$

$y_i - M_\alpha(x_i)$

Split Data (X, y) into

(X_{train}, y_{train}) & (X_{test}, y_{test})

1. Build model α on (X_{train}, y_{train})
2. Evaluate α on (X_{test}, y_{test})
   - choose model

   or

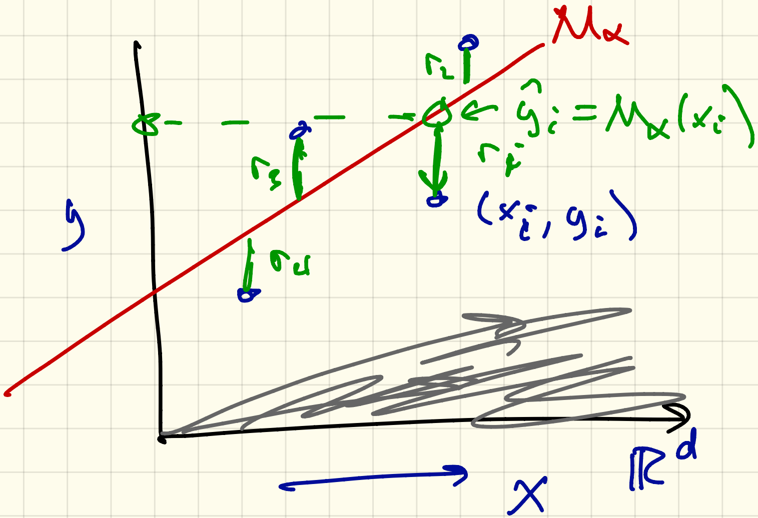
   - predict variance

not both

approx of new data not yet seen.

$$r_i = M_\alpha(x_i) - y_i$$
$$= \hat{y}_i - y_i$$

$$y_i = \hat{y}_i + r_i$$



$$\mathrm{Var}(X, y)M_\alpha = \mathop{E}_i\left[\left(y_i - M_\alpha(x_i)\right)^2\right]$$

$$E[y_i \mid x_i, \alpha]$$

Eigenvectors

define w.r.t. $\boxed{\text{square}}$ matrix $M \in \mathbb{R}^{n \times n}$
(in this course)

start w/ data matrix $A \in \mathbb{R}^{n \times d}$

$$M = AA^T$$

ensures M is $\boxed{\text{positive semi-definite.}}$

ensures for <u>any</u> $x \in \mathbb{R}^n$

$$x M x^T \geq 0$$

square, psd matrix $M \in \mathbb{R}^{n \times n}$

↳ all real, non-negative eigenvalues

Now

$$\boxed{M v = \lambda v}$$

largest eigenvalue λ_1

correspond w/ v_1

in direction of most variance in A

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

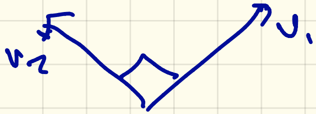
eigenvector

$\lambda \geq 0$

$\lambda \in \mathbb{R}$

eigenvalue

n orthogonal & distinct eigenvectors



eigen vectors v_1, v_2 satisfy $\langle v_1, v_2 \rangle = 0$