Agenda:

1. Feature Transformations
2. Assumptions of linear Reg.
3. Check from Sol.

Feature Transformations.

$\hat{y} = w_0 x_0 + w_1 x_1 + w_2 x_2 + \dots + w_d x_d$

$\text{LogSqrt}^* = \langle x_0, x_1, x_2, \dots, x_d \rangle$

Case I: $\boxed{w_1 = 0}$

Assume f_1 = mileage & $w_1 = 0$

$\hat{y} = w_0 x_0 + w_1 x_1 + \dots + w_d x_d$

Case II: $\boxed{w_1 = +ve}$

f_1 has no effect (unique) on "fare".

$w_1 \to$ "Doesn't matter"

Case III: $\boxed{w_1 = -ve}$

f_1: odometer reading; $w_1 = -ve$

$w_1 = 2$ & $w_2 = 10$

f_1: $w_1 = 2$; $f_2 = w_2^2$ $\dots$

f_1 is more imp.

f_2 is more imp.

$w_1 = -5$ & $w_2 = 10$

f_1: $-5f_1 = -5x_1$ $\dots$

f_2 is more imp.

$w_1 = -8$ & $w_2 = +4$

f_1 $\quad$ f_2

f_2 is more imp. $\boxed{w_1}$

Step 2

$\hat{y} = w_0 + w_1 x_1 + \dots + (+w_2 x_2) + \dots + (-w_d x_d)$ outputs, $\boxed{x_2, x_d}$

$\boxed{x_2}$ $\boxed{x_d}$

$\boxed{\text{Fare}} \to \boxed{\text{Not Valid}}$

$\boxed{\text{Fare}}$

$\boxed{\text{Not Valid}}$

$(+) \to$ Fridge $\quad$ $(-) \to$ Odometer

Robots implications.

Assumptions

Linear Regression

Algebraic Explanation (Geometric) $\quad$ Probabilistic (MLE)

→ Regularizers

1) Transformation of linearity

$y = f_1(x_1)$

2) Multicollinearity (Feature)

Collinearity $\quad$ $f_1 \propto f_2$

So $f_1 = \alpha f_2 + \beta$

$\text{Age} = \frac{\alpha}{2} \text{Dr} + \frac{\text{Dr}+2}{\beta}$

$\therefore F = \boxed{\frac{\alpha}{2}} \text{Dr} + \boxed{\frac{\text{Dr}+2}{\beta}}$

Multicollinearity

Collinearity means matrix fails

f_1, f_2, f_3, f_4

$f_1 = \alpha_1 f_2 + \alpha_2 f_3 + \alpha_3 f_4 + \beta$

Linear Combination

$\boxed{\text{Dr}}$: $\alpha_1 f_2 + \alpha_2 f_3 + \beta$

Why MC is a Problem?

Let $w = \langle 2, 1, 3 \rangle$ $\quad$ (unique)

$w_2 = \langle x_0, x_1, x_2 \rangle$ $\quad$ $\boxed{x_2 = 2x_1}$

$\hat{y}_1 = w_0 x_0 + w_1 x_1 + w_2 x_2 \longrightarrow \hat{y}_1$

$= 1 \cdot x_0 + 2 \cdot [1 \cdot x_1] + 3 \cdot x_2$

$= x_0 + 3x_1 + 3x_2$

$\Rightarrow \langle 1, x_0, x_1 + 3x_2 \rangle \longrightarrow \hat{y}_2$

$\hat{y}_1 = \langle 1, x_0, 3 \rangle$

$w' = \langle 1, 1, 3 \rangle^T$

$w'' = \langle 1, 0, 3 \rangle^T$

$\langle \dots, \dots \rangle^T$

"Length" is not unique.

$|y| \to$ Norm trick to use it for Robust implications.

If $\boxed{\text{MC}}$ $\to$ weights unconstrained

$\downarrow$

result in Robust imp.

3) VIF [Variance Inflation Factor]

$f_1, f_2, f_3, \dots, f_d$

$\to$ Do linear Regression

① $f_2 \leftarrow \langle f_1, f_3, \dots, f_d \rangle$ $\quad$ $(-\infty, \infty)$ $\quad$ R_2

$\langle w_0, w_1, w_3, \dots, w_d \rangle$ $\quad$ $\boxed{\text{Fit}}$

Time Consumed

② $f_3 \leftarrow$ Not checkbox $\langle f_1, f_2, f_4, \dots, f_d \rangle$ $\quad$ R_3

③ f_4

$\boxed{VIF_j = \frac{1}{1-R_j^2}}$

if $R_j^2 \to 1$ ++++ : $VIF_j = \frac{1}{1-1} \Rightarrow \infty$

if $R_j^2 \to 0$: $VIF_j = \frac{1}{1-0} \Rightarrow 1$

if $R_j^2 \to -ve$: $VIF_j = \frac{1}{1-ve} \Rightarrow 0$

$VIF_j \to (1, \infty)$ → Multicollinearity

$\to (1, 0)$ → Linearity

[Rule of Thumb]

$VIF_j \geq 10$: High MC

$VIF_j \geq 5$: Mid MC

$f_2 = f_1$

$f_2 = \alpha f_1 + \beta$

is more multicollinearity?

$f_2 = \alpha f_1 + \beta$

$\boxed{f_1 = \log(f_2) + \beta}$

Non linear fn. $\longrightarrow$ NO MC.

3) Residuals are normally distributed

$y = f_{\theta}(x) + \epsilon$ $\quad$ $N(0, \sigma)$

$\boxed{y - \hat{y}} = \epsilon$

residual

Fig.: $\epsilon = y - \hat{y}$

Mid σ^2

Low σ^2

Mid σ^2

Error → 0 → high value

Linear Regression

"Robust"

Independence

$\sigma_{\epsilon_i, \epsilon_j} = \frac{1}{n}(y_i - \hat{y}_i)^2$

4) Heteroskedasticity should not exists.

Homoscedasticity

$\epsilon_i \sim N(0, \sigma^2)$ $\quad$ $\boxed{\text{Same}}$ $\boxed{\text{Mean}}$

$\boxed{\text{Same}}$

$\boxed{\text{Same}}$

$\boxed{\text{Same}}$

5) No Autocorrelation

$\langle x, y \rangle$ $\quad$ $\boxed{\text{Error Data Points}}$

Any Data Point should be independent function.

$\boxed{\text{Fn}}$ ① $\boxed{\text{Trained}}$ $\boxed{\text{Data}}$ → $\boxed{\text{Model}}$

$\boxed{\text{Fn}}$ ② $\boxed{\text{Trained}}$ $\boxed{\text{Data}}$ → $\boxed{\text{Model}}$

(error) $\quad$ $\boxed{\text{Error}}$ $\geq \epsilon$

Check from Sol / Normal Eq.

Normal Eq.: Analytically Solve "w"

$\frac{\partial J}{\partial w} = 0$

$\boxed{w^T = (X^T X)^{-1} X^T y}$

$(d \times 1)$ $\quad$ $(d \times n)(n \times d)^{-1}$ $\quad$ $(d \times n)(n \times 1)$

$(d \times d)$

$(d \times 1)$

$(d \times 1)$

$\Rightarrow$ Error function ⊖ $[y^T - \hat{y}^T]^2$

$\Rightarrow [y^T - \hat{y}^T]^2$

Error = $[y - Xw]^2$ $\quad$ $\boxed{\hat{y} = Xw}$

$[y - Xw]^T \cdot [y - Xw]$

$(y^T - (Xw)^T) \cdot [y - Xw]$ $\quad$ $(ABC)^T = C^T B^T A^T$

$[y^T - w^T X^T] \cdot [y - Xw]$

$\boxed{\text{Error} \Rightarrow y^T y - y^T X w - w^T X^T y + w^T X^T X w}$ $\quad$ Error is a scalar.

$\left[\frac{\partial \text{Error}}{\partial w} = 0 \right]$

$\frac{\partial}{\partial w}$ $\quad$ $\boxed{\text{Partial derivative}}$

$\Rightarrow -2X^T y + 2X^T X w = 0$

$\Rightarrow 2X^T X w = 2X^T y$

$\Rightarrow \boxed{w^T = (X^T X)^{-1} X^T y}$