

$$f(x; \beta) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \dots + \beta_p x^p$$

→ polynomial representation, or basis expansion

→ simple conceptually
→ " computationally

β is a $(p+1)$ -dimensional vector of regression coefficients

$$f(x) = X\beta$$

(n × 1) (p+1) × 1

$$X = \begin{pmatrix} 1 & x_1 & x_1^2 & \dots & x_1^p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^p \end{pmatrix} \leftarrow \text{design matrix}$$

We may have more than one covariates

$$f(x_1, x_2; \beta) = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \beta_4 x_2^2 + \dots$$

$$f(x_1, x_2; \beta) = \beta_0 + \beta_1 x_1 + \beta_2 x_1/x_2 + \beta_3 x_2^2 + \dots$$

X does not need to be continuous (for example, type-of-fuel in our data is binary, gasoline/diesel)

$$x_5 = \begin{cases} 0 & \text{if } x_2 \text{ is gasoline} \\ 1 & \text{if } x_2 \text{ is diesel} \end{cases} \leftarrow \text{dummy variable}$$

β_5 in the R-code : expected difference in km per liter when we change from gasoline to diesel given all other variables equal

when there are more modalities $x_5 = \{\text{gasoline, diesel, electric}\}$

$$x_{51} = \begin{cases} 0 & \text{otherwise} \\ 1 & \text{if diesel} \end{cases}$$

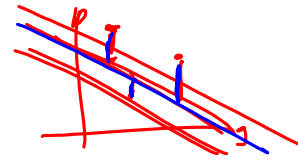
$$x_{52} = \begin{cases} 0 & \text{if otherwise} \\ 1 & \text{if electric} \end{cases}$$

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{matrix} \rightarrow \text{electric} \\ \rightarrow \text{gasoline} \\ \rightarrow \text{diesel} \\ \rightarrow \text{electric} \end{matrix}$$

No matter of the nature of X , we need to estimate β

$$\hat{\beta} = \arg \min_{\beta} \{RSS(\beta)\} = \arg \min_{\beta} \left\{ \sum_{i=1}^n (y_i - \hat{f}(x_i; \beta))^2 \right\}$$

$$\hat{\beta} = \arg \min_{\beta} \{ (y - X\beta)^T (y - X\beta) \}$$



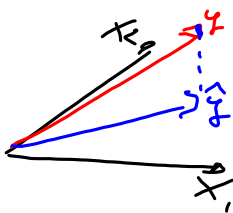
$$RSS(\beta) = (y - X\beta)^T (y - X\beta)$$

$$\frac{\partial RSS(\beta)}{\partial \beta} = -2X^T(y - X\beta) = 0$$

$$-X^T y + X^T X \beta = 0$$

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

least squares estimator



$$\hat{y} = X\hat{\beta}$$

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

P projection matrix

$$\varepsilon_i \sim \mathcal{N}(0; \sigma^2)$$

Using the $RSS(\hat{\beta})$, we can also estimate σ^2

$$s^2 = \frac{RSS(\hat{\beta})}{n - p - 1} = \frac{\sum_{i=1}^n (y_i - \hat{f}(x_i; \hat{\beta}))^2}{n - p - 1}$$

and the variance of $\hat{\beta}$

$$\text{Var}(\hat{\beta}) = s^2 (X^T X)^{-1}$$

Assuming $\varepsilon \sim \mathcal{N}(0; \sigma^2)$, $\hat{\beta} \sim \mathcal{N}(\beta; (X^T X)^{-1} \sigma^2)$

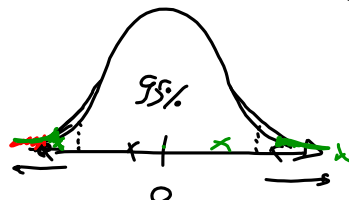
linear Gaussian regression

$$t\text{-values } t_j = \frac{\hat{\beta}_j}{\text{se}(\hat{\beta}_j)} = \frac{\hat{\beta}_j}{\sqrt{s^2 (X^T X)^{-1}_{jj}}}$$

$\hat{\beta}_j$

$$\begin{matrix} X^T X & \text{matrix} \\ p \times n & n \times p \\ \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \end{matrix}$$

$$t \sim t_{n-p-1}$$



p-value: probability of obtaining a t-value that is larger in absolute value of the one actually obtained, under the null hypothesis $\beta_j = 0$

$$y = \beta_0 + \beta_1[\text{engine-size}] + \beta_2[\text{engine-size}^2] + \beta_3[\text{engine-size}] + \beta_4[\text{type of fuel} = \text{diesel}] + \varepsilon$$

To evaluate the goodness-of-fit, we can use the coefficient of determination R^2

$$R^2 = 1 - \frac{RSS(\beta)}{TSS\beta} = \frac{\text{how much the model explains (in terms of deviance)}}{\text{total deviance}}$$

$$= 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad \text{where } \bar{y} \text{ is } \frac{1}{n} \sum_{i=1}^n y_i$$

$0 \leq R^2 \leq 1$ • R^2 is easy to interpret, although it is not linear
• not oversimplify

→ graphical diagnosis of the model fit based on the residuals

NOTATION		
X	matrices	
x	scalars or vectors	numel scalar bold vector
β	parameter	$i = 1, \dots, n$ observations
$\hat{\beta}$	estimator	$j = 1, \dots, p$ variables
$\hat{\beta}$	estimate	

$$\hat{\epsilon}_i = y_i - \hat{y}_i$$

$\hat{\epsilon}_i$ are surrogates of ϵ_i , which cannot be observed

scatter
→ plot of the residuals

→ average 0, violation of homoscedasticity (same σ^2)

→ quantile-quantile plot

y -axis = standardized $\hat{\epsilon}_i$
 x -axis = expected values under normality

→ in our example the points do not lie on the diagonal, as they should.

Sometimes one needs a test for a group of coefficients.

We can use the F-statistic

$$F = \frac{(RSS_0 - RSS_1) / (p_1 - p_0)}{RSS_1 / (n - p_1 - 1)}$$

$$N F_{p_1 - p_0, n - p_1 - 1}$$

where,

RSS_1 is the residual sum of squares of the larger model

p_1 is the number of parameters of the larger model

RSS_0 is the residual sum of squares of the sparser model

p_0 is the number of parameters of the sparser model



Transformation of the response

linear model $\rightarrow$ linear in the parameter

we can also transform the response

$$\underline{\underline{x_3 = x_1/x_2}}$$

E.g. : transform distance-per-liter

$$\text{consumption} = \frac{1}{\text{distance-per-liter}}$$

$$\text{consumption} = \beta_0 + \beta_1 \text{engine-size} + \beta_2 \text{type-of-fuel} + \varepsilon$$

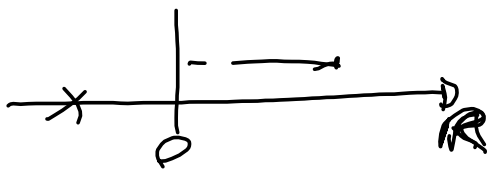
$$\text{distance-per-liter} = \beta_0 + \beta_1 \text{engine-size} + \beta_2 \text{engine-size}^2 + \beta_3 \text{engine-size}^3 + \beta_4 \text{type-of-fuel}$$

$$\text{consumption} = \beta_0 + \beta_1 \text{engine-size} + \beta_2 \text{type-of-fuel}$$

$$\hat{y} = \text{consumption}$$

$$\frac{1}{\hat{y}} = \text{distance-per-liter}$$

More often, one use the logarithmic transformation if the variables (response and covariates) are positive by definition



$$y > 0$$

$$\log(y) \quad -\infty < \log(y) < +\infty$$

$$\underline{\log(\text{distance-per-liter})} = \beta_0 + \beta_1 \log(\text{engine-size}) + \beta_2 \text{type-of-fuel} + \varepsilon$$