Estadística III para Ingenieros de Sistemas

Jose Daniel Ramirez Soto 2023
jdr2162@columbia.edu

Agenda

- anuncios varios
    - https://forms.office.com/r/LeFfxyg4rQ
- modelos de analitica (machine learning-ML) Supervisado
    - Regresión
- Práctica de regresión en Python

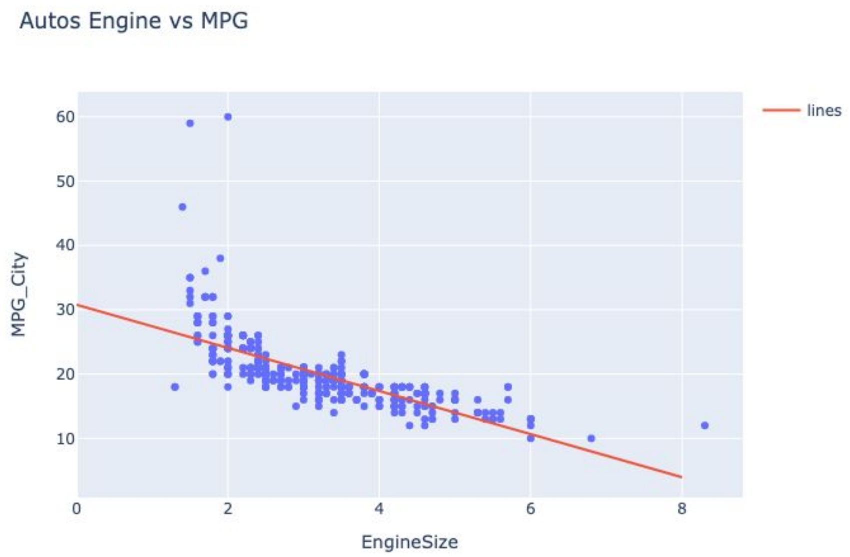
Supervisado, Regresión ejemplo

Tomando los datos de los carros, vamos a crear una regresión utilizando el tamaño del motor.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x,$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x},$$



Medidas de error en los problemas de regresión:

$$MSE = \frac{1}{n}\sum_{i=0}^{n}(y_i - \tilde{y}_i)^2, MAPE = \frac{1}{n}\sum_{i=0}^{n}\frac{|y_i - \tilde{y}_i|}{y_i}$$

Ej $y=20$, $\tilde{y}=19$, mse=1 , mape = 0.05

* A Course of Machine Learning http://ciml.info/

Supervisado, Regresión P-value (menor a 0.005)

Para conocer la relevancia de una variable se utilizan hipótesis test :

null hypothesis H0 :No existe relación entre las variables y el coeficiente es 0

H0 : β1 = 0

alternative hypothesis Ha : Existe relación entre las variables

Ha : β1 ≠ 0,

$$t = \frac{\hat{\beta}_1 - 0}{\mathrm{SE}(\hat{\beta}_1)}, \quad \mathrm{SE}(\hat{\beta}_1)^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \mathrm{SE}(\hat{\beta}_0)^2 = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]$$

* A Course of Machine Learning http://ciml.info/

Supervisado, Regresión ejemplo

MPG_City (y)	EngineSize(X)
18	2.5
18	3.8
21	2.0
29	2.0
18	4.3
25	2.0
13	5.6
16	3.0
16	3.2
29	1.8
17	4.6
16	5.0

TRAIN

TEST

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2},$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x},$$

$$MSE = \frac{1}{n}\sum_{i=0}^{n}(y_i - \tilde{y}_i)^2, MAPE = \frac{1}{n}\sum_{i=0}^{n}\frac{|y_i - \tilde{y}_i|}{y_i}$$

x̄ = 30.2/10 = 3.2 , ȳ = 203/10 = 20.3

β1_num = (2.5 - 3.2)(18 -20.3)+ (3.8 - 3.2)(18 -20.3) ….. + (1.8 - 3.2)(29 -20.3) = **-48.06**

β1_den = (2.5- 3.2)^2+ (3.8 - 3.2)^2 +….+ (1.8 - 3.2)^2 = **13.816**

β1= **-48.06/13.816 = -3.47**

β0 = 20.3-(-3.47)*3.2 = **31.404**

ŷ11 = 4.6*(-3.47) + 31.404 = 15.442
ŷ12 = 5.0*(-3.47) + 31.404 = 14.054

MSE Train = 11. = ~ 3.41

MSE = ((17-15.44)^2 + (16-14.054)^2)/2 = 3.110258 , el error es (3.11)^0.5 **= 1.76**
MAPE = (ABS(17-15.44)/17 +abs(16-14.054)/16)/2 = **0.10**

* A Course of Machine Learning http://ciml.info/

Supervisado, Regresión P-value (menor a 0.005) ejemplo

Universidad Católica de Oriente

Para conocer la relevancia de una variable se utilizan hipótesis test :

null hypothesis H0 :No existe relación entre las variables y el coeficiente es 0

H0 : β1 = 0

alternative hypothesis Ha : Existe relación entre las variables

Ha : β1 /= 0,

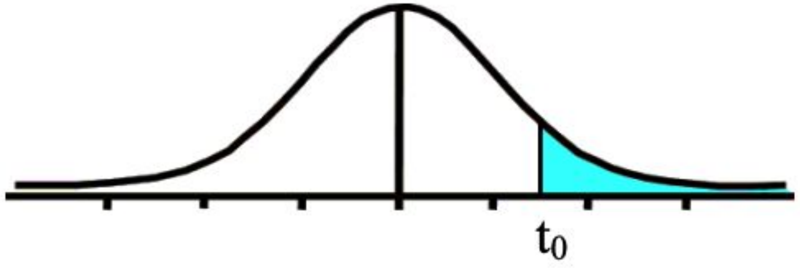
$$t = \frac{\hat{\beta}_1 - 0}{\mathrm{SE}(\hat{\beta}_1)}, \ \mathrm{SE}(\hat{\beta}_1)^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \mathrm{SE}(\hat{\beta}_0)^2 = \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]$$

σ^2= (y-ŷ) = ((18-22.7)^2+ (18-18.18)^2…+(29-25.14)^2) = 116.50/10 =11.65
SE(β1) = σ^2/(x-x̄)^2 = 116.50/((2.5- 3.2)^2+ (3.8 - 3.2)^2 +….+ (1.8 - 3.2)^2) = 11.65/13.8 = 0.84
t = (-3.47 - 0)/0.84^0.5 **= -3.7860**
grados de libertad son el numero de observaciones -1 - el numero de parámetros que es el numero de betas.

* A Course of Machine Learning http://ciml.info/

Supervisado, Regresión P-value (menor a 0.005) ejemplo

Tabla t-Student



Grados de libertad	0.25	0.1	0.05	0.025	0.01	0.005
1	1.0000	3.0777	6.3137	12.7062	31.8210	63.6559
2	0.8165	1.8856	2.9200	4.3027	6.9645	9.9250
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8408
4	0.7407	1.5332	2.1318	2.7765	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9979	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693

* A Course of Machine Learning http://ciml.info/

https://cms.dm.uba.ar/academico/materias/1ercuat2015/probabilidades_y_estadistica_C/tabla_tstudent.pdf

Supervisado, Regresión

MPG_City (y)	EngineSize(X)
18	2.5
18	3.8
21	2.0
29	2.0
18	4.3
25	2.0
13	5.6
16	3.0
16	3.2
29	1.8
17	4.6
16	5.0

* A Course of Machine Learning http://ciml.info/

Calcular el R-squared utilizando la regresión del ejemplo.

β_1= **-48.06/13.816 = -3.47**

β_0 = 20.3-(-3.47)*3.2 = **31.404**

ŷ = x* β_1 **+**β_0

$$R^2 = \frac{\mathrm{TSS} - \mathrm{RSS}}{\mathrm{TSS}} = 1 - \frac{\mathrm{RSS}}{\mathrm{TSS}}$$

$$\mathrm{RSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2.$$

$$\mathrm{TSS} = \sum (y_i - \bar{y})^2$$

Supervisado, Regresión

Resultado de una regresión, R-squared es la proporción de la varianza de las millas por galón que es explicada por el motor .

OLS Regression Results

Dep. Variable:	MPG_City	**R-squared:**	0.503
Model:	OLS	**Adj. R-squared:**	0.502
Method:	Least Squares	**F-statistic:**	431.7
Date:	Thu, 16 Mar 2023	**Prob (F-statistic):**	9.86e-67
Time:	17:17:56	**Log-Likelihood:**	-1165.8
No. Observations:	428	**AIC:**	2336.
Df Residuals:	426	**BIC:**	2344.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	30.7772	0.546	56.388	0.000	29.704	31.850
EngineSize	-3.3523	0.161	-20.778	0.000	-3.669	-3.035

Omnibus:	447.615	**Durbin-Watson:**	1.310
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	25957.663
Skew:	4.519	**Prob(JB):**	0.00
Kurtosis:	40.066	**Cond. No.**	11.1

$$R^2 = \frac{\mathrm{TSS} - \mathrm{RSS}}{\mathrm{TSS}} = 1 - \frac{\mathrm{RSS}}{\mathrm{TSS}}$$

$$\mathrm{RSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2.$$

$$\mathrm{TSS} = \sum (y_i - \bar{y})^2$$

* A Course of Machine Learning http://ciml.info/

Supervisado, Regresión con más variables

Calcular Betas o w para muchas variables

$$\hat{y} = w^T\mathbf{x} + b = \sum_{i=1}^{p} w_i x_i + b$$

$$\min_{\boldsymbol{w}} \quad \mathcal{L}(\boldsymbol{w}) = \frac{1}{2}\,||\mathbf{X}\boldsymbol{w} - \boldsymbol{Y}||^2 + \frac{\lambda}{2}\,||\boldsymbol{w}||^2$$

$$\boldsymbol{w} = \left(\mathbf{X}^\top\mathbf{X} + \lambda\mathbf{I}_D\right)^{-1}\mathbf{X}^\top\boldsymbol{Y}$$

* A Course of Machine Learning http://ciml.info/

Supervisado, Regresión

La predicción es el valor de la función f(x) con los nuevos dato $\hat{y}$

$$\hat{y} = w^T\mathbf{x} + b = \sum_{i=1}^{p} w_i x_i + b \quad = \quad \underbrace{\begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,D} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,D} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N,1} & x_{N,2} & \cdots & x_{N,D} \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_D \end{bmatrix}}_{\boldsymbol{w}} = \underbrace{\begin{bmatrix} \sum_d x_{1,d} w_d \\ \sum_d x_{2,d} w_d \\ \vdots \\ \sum_d x_{N,d} w_d \end{bmatrix}}_{\hat{\mathbf{Y}}} \approx \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}}_{\hat{\mathbf{Y}}}$$

* A Course of Machine Learning http://ciml.info/