Lec : 04
01.29.24

"Linear Regression"

LR → simple approach to
↓
supervised learning

It makes some assumptⁿ:

i) All predictors are Independent
↓
each of the features will affect independently

ii) Y → the label / target depends on the predictors.
(There is a relationship between X and Y)

iii) The relationship is linear

The form of LR → equation of a line

$$\boxed{y = c + mx}$$

↓
this form we assume for $f(x)$

$$Y = \underbrace{\underbrace{\beta_0}_{\text{intercept}} + \underbrace{\beta_1}_{\text{slope}}}_{\text{coefficient / parameters}} x + \underbrace{\epsilon}_{\text{error terms}}$$

β_0 and β_1 → We estimate for those values
↓
so that output can be predicted

Linear reg assumes = Linear funct^n (Straight line)

$$y = c + mx$$

$$\Rightarrow y = f(x)$$

$$\Rightarrow y = \beta_0 + \beta_1 x + \varepsilon$$

Estimate $\hat{\beta}_0$ and $\hat{\beta}_1$ to predict $\hat{y}$

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

↓

need to find these values.

$\hat{\beta}_0 = ?$ $\hat{\beta}_1 = ?$

↓

Least square method → efficient method for finding the values of β_0 and β_1.

$$y = f(x)$$

$$\hat{y} = \hat{\beta}_0 + \beta_1 X$$

$$e_1^2 = (y - \hat{y})^2 =$$

e_2^2, e_3^2, e_4^2, ... e_n^2

All of these errors together:

Residual Sum Square

$$(RSS) = e_1^2 + e_2^2 + \dots + e_n^2$$

$$(y_1 - \hat{\beta}_0 + \hat{\beta}_1 X)^2$$

$\hat{\beta}_0 = ?$ $\hat{\beta}_1 = ?$ what are the estimates?

→ or finding

$$RSS = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

Partial derivative of RSS, with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$, and set it equal, zero.

$$\boxed{\hat{\beta}_0 = \bar{y} - \beta_1 \bar{x}}$$

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

Where, $\bar{x}, \bar{y}$ = sample means of X and Y

$$f(x) = \hat{\beta}_0 + \hat{\beta}_1 x$$

Evaluation of Linear regression:

i) Standard Errors : will give us values
→ How close $\hat{\beta}_0$, $\hat{\beta}_1$ estimates to the true value → 0 to ∞
→ the lower, the better

confidence intervals :

[±2, ±10]

95% sure that we are in the right track. → true value will be inside this range.

R^2 = proportion of variance explained.
↓
(Accuracy equivalence) ~~Accuracy equivalence~~
How much explanatⁿ the model can give us.?

[0, 1]

[0%, 100%]

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

TSS = Total sum squared

$= \sum_{i=1}^{n} (y_i - \bar{y})^2$

Hypothesis: Assumptⁿ that needs to be tested.

null hypothesis → no relation

alternative " → some "

null hypothesis

↓

train and test a model

↓

Accept / reject based on this model

ex: $f(x) = \beta_0 + (\beta_1 x) \rightarrow$ if $\beta_1 = 0$

$= \beta_0$

t-statistic → way to reject / accept the hypothesis

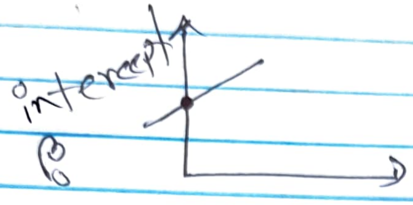
↓

Some form of relationship

↓

which coefficient has relation individually.

f-statistic: At least 1 co-efficient has the relationship



$$f(x) = \hat{\beta}_0 + \beta_1 x$$

$$= 7.0325 + 0.0475(x)$$

P-value $\rightarrow$ $\boxed{p < 0.05}$

All the results we have shows ~~is~~ statistically significant coefficient.

Lec: 05
01.31.23

"coding session"

writing code to implement Lear regression

↓

data

Datasets in ISLP

↓

Boston data

↓

Simple LR:
just 1

$$f(x) = \beta_0 x_0 + \beta_1 x_1$$

$\beta_0 \times 0$

$\beta_0 \times 1$

result: lstat : −0.9500
(negative relationship)

lstat: lower status of the populatn

$t \rightarrow -24.528$

MLR

x_0	x_1	x_2	y (model)

positive relatn:

$\boxed{\text{lstat : age}}$

doesn't matter how old

At first without interactn → individually → then together qualitative - data:

0 = No
1 = yes

$f(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$

$$\frac{1}{n}\sum_{i=1}^{n} I(y_i \neq \hat{y}_i)$$

$y_i \neq \hat{y} \rightarrow 1$

$y_i = \hat{y} \rightarrow 0$

$$\text{Ave}(I(y_0 \neq \hat{y}_0))$$

$\Pr(Y = j \mid X = x_0) = \frac{1}{K}\sum_{i \in N_0} I(y_i = j)$

$$CV_{(n)} = \frac{1}{n}\sum_{i=1}^{n} \text{Err}_i$$

1.6 2.4 8.8

1.6 M
2.4 K
8.7 O

Lec : 06
02-05-24

(i) Is there a relationship between X and Y?
→ t-statistics / f-statistic
whether / not reject null hypothesis

(ii) How strong is the relationship between X and Y?
co-efficient > impact
→ R^2 (0 - 1 → ↑↑ → better)

(iii) Which of the predictors are associated with y?
→ P-value, < 0.05

(iv) Is there synergy among the predictors?
→ P-value, < 0.05
examine the " associated with each Predictor' (t-stat)

(v) How accurate is the model for future predictⁿ?
→ co-efficient within the
→ $\boxed{\text{confidence-interval}}$

classificat^n:

predicting classificat^n problem

feature vector $X \to Y$

qualitative response

no quantitative

Try to predict the probability of label y.

eg: ~~P(x)~~

$P(X \mid brown) =? \; 0.5$ ✓

$P(X \mid blue) =? \; 0.2$

$P(X \mid green) =? \; 0.3$

⋮

$P(X \mid y_n) = ?$

income	balance	default
–	–	No
–	–	No
–	–	

↓

co-efficient of balance needs to be ↑↑ but lower in income

Linear regression

Why not LGR model?

$\hat{y} > 0.5 \rightarrow 1$ yes

else

$\hat{y} < 0.5 \rightarrow 0$ NO

But LR → predicts a lot and can go above 1 or below 0.

Logistic Regression:

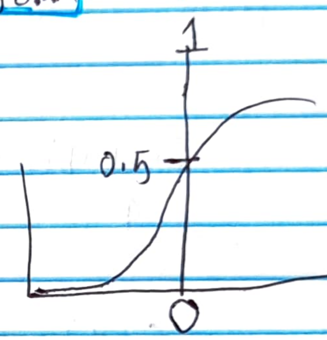
→ Parametric method

→ ** Data of a certain form / functional form **

$y = c + mx$

$y = \beta_0 + \beta_1 x$

→ Assume sigmoid funct^n: ≈ [0, 1]



above 0.5 ⇒ yes

below → NO

$$S(x) = \frac{1}{1 + e^{-x}}$$

Euler constant 2.71

$$= \frac{1}{1 + (2.71)^{-x}}$$

any value we provide, will be in the range of [0 - 1]

$$P(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$

[value within 0 and 1

$$1+e^{-x+x+y+x}$$

X_0	x_1	y_1	
			$e^{\beta_0+\beta_1 X}$ ↓ we are gonna predict β_0 and β_1 co-efficient

We have funct^n :

$$P(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

How do we find this estimates ;

* Maximum Likelihood : find the Probability of predictor giving y/label = 1.

$S(x) = \frac{1}{1+e^{-x}} = \frac{1}{1+\frac{1}{e^x}}$

$$= \frac{1}{\cancel{1}} \times \frac{1}{\cancel{e^x}}$$

$$=$$

$$\underbrace{\pi_1\, p(x_i)}_{\text{likelihood of predicting } 1} \cdot \underbrace{(1-p(x_i))}_{\text{and } 0}$$

↓

Predictor will tell which one is more likely

Maxi likelihood for each of the predictor

f - statistic → overall relationship (1 predictor has ?)
t - statistic → individual relationship → individual predictor

Z → (individual predictor relationship)

Making Pred :

$$\hat{p}(y) = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 X}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1 X}}$$

$$\underbrace{\hat{p}(x) = \hat{p}(x)}_{\text{comparing whether they are same/not}}$$

g) ↑↑ balance → predicting good.
co-efficient
0.4049 → can be any number
Stu(yes) → has ↑↑ magnitude than the balance predictor.

multiple predictors:

slide 11 →

Student → neg. coefficient

↓

not a good sign which is fooling the model

opposite prediction of model.