BUSS6002 Data Science in Business

UNIVERSITY OF SYDNEY · POSTGRADUATE COURSEWORK · SEMESTER 2, 2026

Full guide + ask Sia about BUSS6002

SIDE 1/2 PRIORITY-ORDER PLAN FOR YOUR HANDWRITTEN A4 — **HANDWRITTEN A4 PLANNER** — Methods · formulas · decisions · assumptions

01 · Exam map

COURSE DATA + FINAL INSTRUCTIONS

Mid-semester | Individual | Final exam 45%

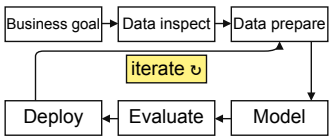
FINAL COMPONENT	MARKS	FAST IMPLICATION
MCQ	12	definitions, traps, classify
Short answer	27	show formula + working
Python code	6	runnable syntax from memory

Material permitted: One A4 sheet of handwritten notes DOUBLE SIDED. "Handwritten" means that the notes must be written directly by hand on the physical sheet of paper that you bring into the exam. (BUSS6002 final exam instructions and format (Semester 1 2026; confirm for Semester 2 on Canvas))

PRIORITY: *Copy formulas and decision triggers first; then notation traps; then the recall tables. Do not spend scarce handwriting space on introductions.*

02 · Process + easy marks

CH01–03



CRISP-DM in order; iteration is allowed, but each phase has a distinct job.

TRIGGER	ANSWER
What / why / next / action?	descriptive / diagnostic / predictive / prescriptive
Big data	Volume · Variety · Velocity · Veracity; generation mechanism, not bytes
Decision context	simple / complicated / complex / chaotic: name cause–effect signature
Missingness	MCAR: none; MAR: observed driver; MNAR: missing value itself

CRISP-DM PHASE	MARK-WINNING JOB
Business understanding	problem, success, stakeholders, cost/ethics
Data understanding	collect, describe, explore, verify quality
Data preparation	clean, integrate, derive features
Modeling	choose, fit and tune method
Evaluation	check technical + business success
Deployment	deliver, monitor, maintain

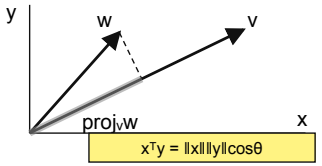
EDA ANCHORS

```
IQR = Q3 - Q1; fences = [Q1 - 1.5·IQR, Q3 + 1.5·IQR]
cor(x,y)=cov(x,y)/[sd(x)sd(y)] ∈ [-1,1]
```

cor = 0 does not imply independence. A circled box-plot point is extreme, not automatically erroneous. Investigate before deleting.

03 · Vectors + matrices

CH04 + CH06



Source redraw: dot product links algebra to angle; orthogonal means $x^Ty=0$.

VECTOR CHAIN

$(x,y) = \Sigma x_i y_i = x^Ty \to$ scalar

$\|x\|_2 = \sqrt{(x^Tx)};\ \|x\|_2^2 = x^Tx$

$d(x,y)=\|x-y\|_2;\ x \perp y \Leftrightarrow x^Ty=0$

OPERATION	EXAM CHECK
ax+by	equal lengths; scale each, then add
linear dependence	find coefficients not all zero giving $\Sigma a_i x_i=0$
AB	(m×n)(n×p)→m×p; row·column
transpose	$(A^T)_{ij}=A_{ji}$; generally AB≠BA

STATISTICS AS INNER PRODUCTS

$\hat{\mu}_x=((1/n)1,x)$

$\tilde{o}(x)=(1/\sqrt{n})\|x\|_2$ (zero-mean)

$co\hat{v}(x,y)=((1/n)x,y)$ (zero-mean)

MATRIX OBJECT	RULE / ROLE
I	AI=IA=A
A^{-1}	$AA^{-1}=I$ when inverse exists
det(A)	zero ⇒ singular; no unique inverse
data matrix X	rows observations; columns variables

HANDWRITTEN NUMPY RECALL

```
x=np.array([3,1,-2]);
y=np.array([2,2,4])
np.dot(x,y)   # one scalar; vectorised
```

Dot product returns one number, not element-wise products. Norm/distance need the final square root; squared norm does not.

04 · K-means

CH05

OBJECTIVE + UPDATE

$WCSS = \Sigma_i \Sigma_{x \in C_i} \|x-\mu_i\|_2^2$

$\mu_i = (1/|C_i|)\Sigma_{x \in C_i} x$

STEP	DO
0	choose k; initialise k centres
1 assign	send each x to nearest μ by squared distance
2 update	replace each μ by mean of new members
3 repeat	until centres stop moving

WCSS is non-increasing and bounded below by 0, hence convergence—but only to a local optimum. Run several starts; keep lowest WCSS. K-means does not choose k; use elbow/business actionability.

SEGMENTATION BASIS	EXAMPLES
demographic	age, income
geographic	location
psychographic	values, lifestyle
behavioural	usage, loyalty

05 · OLS pipeline

CH06

MODEL + ESTIMATOR

$y = X\beta + \varepsilon;\ \hat{y} = X\hat{\beta};\ e = y-\hat{y}$

$\hat{\beta} = \arg\min_\beta \|y-X\beta\|_2^2 = (X^TX)^{-1}X^Ty$

SIMPLE REGRESSION SHORTCUT

$S_{xx}=\Sigma(x_i-\bar{x})^2;\ S_{xy}=\Sigma(x_i-\bar{x})(y_i-\bar{y})$

$\hat{\beta}_1=S_{xy}/S_{xx};\ \hat{\beta}_0=\bar{y}-\hat{\beta}_1\bar{x}$

FIT

$RSS=\Sigma e_i^2;\ TSS=\Sigma(y_i-\bar{y})^2$

$R^2=1-RSS/TSS=(TSS-RSS)/TSS$

ASSUMPTION / READING	WHERE IT BITES
$E(\varepsilon\|x)=0$	unbiased conditional mean; curve signals failure
$var(\varepsilon\|x)=\sigma^2$	constant variance; funnel signals failure
ε uncorrelated with x	estimation validity
Normal ε	needed for t/F inference, not for fitting $\hat{\beta}$

INFERENCE WHEN REQUESTED

$SE(\hat{\beta}_1)^2=\sigma^2/S_{xx};\ t=\hat{\beta}_1/SE(\hat{\beta}_1)$

approx. 95% CI: $\hat{\beta}_1 \pm 2\cdot SE(\hat{\beta}_1)$

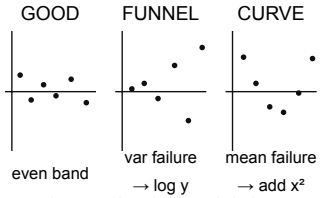
QUESTION GIVES	FAST ROUTE
X and y matrices	normal equation $(X^TX)^{-1}X^Ty$
$\bar{x},\bar{y},S_{xx},S_{xy}$	slope then intercept shortcut
fitted values	e=y−ŷ → RSS → R^2
residual plot	shape → assumption → remedy

Existence of $(X^TX)^{-1}$ requires full column rank: redundant predictors make the normal equation singular. Centre the interpretation on units and the held-fixed predictors.

Interpret $\hat{\beta}_j$ as the average change in y associated with +1 in x_j, holding other predictors fixed—never automatically causal.

06 · Residual decision

CH06 FIG 3



Source redraw: name the visible shape, the broken assumption and one remedy.

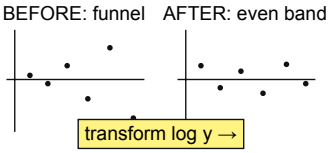
PLOT	DIAGNOSIS	MOVE
flat/even	adequate	no structure left
funnel	heteroscedastic	transform y, e.g. log y
curve/U	missed nonlinearity	add x^2 / basis term

Residual e_i is the vertical observed-minus-fitted gap. RSS squares and sums those gaps. The least-squares line passes through $(\bar{x},\bar{y})$. A high R^2 describes in-sample fit only; it neither proves assumptions nor causality.

07 · Features + ethics

CH07

SYMPTOM / CASE	USE / PRINCIPLE
positive right-skewed y; funnel	model log(y): symmetry + variance stabilisation
curvature	polynomial $x^2,x^3,\ldots$
effect of x_1 depends on x_2	interaction x_1x_2
consent to collect?	Ownership
told how used?	Transparency
PII exposed/re-identifiable?	Privacy
purpose?	Intention
group harm?	Outcomes / disparate impact



Source redraw: a log-response transform can compress the tail and stabilise variance.

BACK-TRANSFORM CORRECTLY

$\log(y)=x^T\beta+\varepsilon$

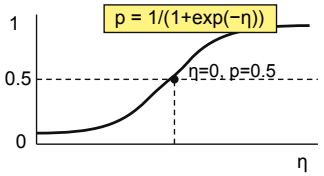
$E(y|x)=\exp(x^T\beta)\cdot E[\exp(\varepsilon)]$

$\hat{y}_{corr}=\exp(x^T\hat{\beta})\cdot(1/n)\Sigma\exp(e_i)$

By Jensen, $E[\exp(\varepsilon)]>1$ for non-degenerate mean-zero error. Thus naive $\exp(x^T\hat{\beta})$ is downward biased; the smearing factor scales it up.

08 · Logistic regression

CH08 FIG 1



Source redraw: η=$x^T\beta$ is log-odds; the sigmoid alone maps it into (0,1).

LINK → ACTIVATION → CLASS

$\log[p/(1-p)] = \eta = x^T\beta$

$\hat{p}=1/[1+\exp(-\eta)];\ \hat{y}=1$ if $\hat{p}>\tau$ (default $\tau=0.5$)

+1 in x_j multiplies odds by $\exp(\beta_j)$

GLM PART	LINEAR	LOGISTIC
response	continuous	Bernoulli
link g	identity	logit
activation g^{-1}	identity	sigmoid
fit	OLS; closed form	MLE; iterative

A probabilistic classifier retains $\hat{p}$; thresholding discards confidence to create a label. BUSS6002 fixes $\tau=0.5$ rather than asking you to optimise it.

Fit by Bernoulli maximum likelihood, not OLS. Positive η ⇒ $\hat{p}>0.5$. Lower τ raises both recall and FPR.

09 · Classifier metrics

CH08

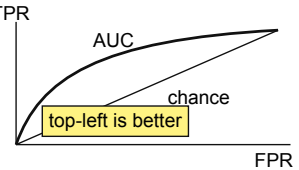
	ACTUAL +	ACTUAL −
PRED +	TP	FP · Type I
PRED −	FN · Type II	TN

READ MARGINS, NOT MEMORY

```
Accuracy=(TP+TN)/N
Recall/TPR=TP/(TP+FN)
Specificity/TNR=TN/(TN+FP)
FPR=FP/(FP+TN)=1-TNR
Precision=TP/(TP+FP)
F1=2TP/(2TP+FP+FN)
```

BUSINESS QUESTION	METRIC
Of real positives, caught?	recall
Of flags, trustworthy?	precision
Both matter, one score?	F1
Rare positives?	do not rely on accuracy
ranking across thresholds?	ROC/AUC

Propensity predicts natural action; campaign-response predicts action given exposure; uplift estimates the causal incremental effect using treatment vs control.

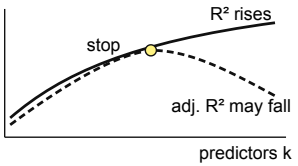


Source redraw: ROC plots TPR against FPR as τ moves; AUC summarises ranking.

AUC near 1 means better ranking; 0.5 is chance. ROC runs from (0,0) to (1,1) and is monotone non-decreasing as the threshold relaxes.

10 · Select complexity

CH09 FIG 1



Source redraw: in-sample fit rewards extra terms; honest selection needs held-out data.

GENERALISATION TARGET

$EPE=E[(y-\hat{y})^2]=\sigma^2 + Bias^2 + Variance$

complexity ↑ ⇒ bias ↓, variance ↑ ⇒ EPE U-shaped

FLEXIBLE BUT LINEAR IN B

$y=\varphi(x)^T\beta+\varepsilon;\ \hat{\beta}=(\Phi^T\Phi)^{-1}\Phi^Ty$

polynomial: $\varphi_i(x)=x^i$

radial: $\varphi_i(x)=\exp[-(x-c_i)^2/(2s^2)]$

DATA SPLIT	ONLY JOB
train	fit each candidate
validation	select lowest MSE
test	one final unbiased report

11 · MLE + optimisation

CH10 FIG 3

LIKELIHOOD

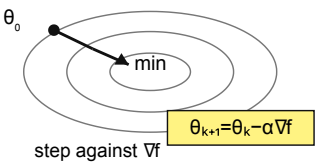
$L(\theta)=p(data|\theta)=\prod_i p(y_i|\theta)$

$\ell(\theta)=\log L(\theta)=\Sigma_i \log p(y_i|\theta)$

$\hat{\theta}=\arg\max_\theta \ell(\theta)$

MLE LINE	MARK
write product likelihood	independence
take logs	product→sum
differentiate	score
set =0; solve	candidate $\hat{\theta}$
second derivative <0	maximum check

Why log? Avoid underflow; simplify differentiation; log is increasing so argmax unchanged. Drop additive terms without θ.



Source redraw: gradient points uphill; subtract it to minimise $f=-\ell$.

A	FAILURE
too large	overshoot/diverge
too small	very slow

OLS has closed form. Logistic score is nonlinear, so minimise cross-entropy $f=-\ell$ iteratively; $\nabla f(\beta)=X^T(p-y)$.

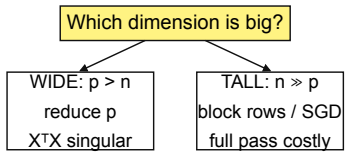
BERNOULLI LOG-LOSS

$f(\beta)=-\Sigma_i[y_i \log p_i+(1-y_i)\log(1-p_i)]$

argmin is the parameter location; ***min*** is the achieved objective value. Independence is what permits L to factor into a product and ℓ into a sum.

12 · Scale by shape

CH11



Shape decides the remedy: large p and large n are different problems.

COMPLEXITY + SGD

$O(1)<O(\log n)<O(n)<O(n \log n)<O(n^2)<O(2^n)$

full gradient: O(n)/step; one-row SGD: O(1)/step in n

Big-O is an asymptotic upper bound on worst-case growth, not an exact operation count. SGD is noisy but updates much more often for tall data.

SHAPE	FAILURE	CORRECT MOVE
p>n	overfit; X^TX singular	feature selection, projection, less flexibility
n≫p	each full pass expensive	blocks or one-row SGD

BUSS6002 Worked drills · traps · fast recall

DATA SCIENCE IN BUSINESS · UNIVERSITY OF SYDNEY · USE BEFORE THE EXAM

Full guide + ask Sia about BUSS6002

SIDE 2/2 | 5 FULL WORKED DRILLS · METHOD CHOOSER · PYTHON RECALL · TOP TRAPS | PRACTISE · CHECK · REPEAT | All numerical values from the course guide

13 · Drill 1: box plot

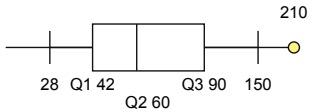
CH03 WORKED EXAMPLE

Given sorted A$ totals: 28, 35, 42, 46, 58, 60, 75, 86, 90, 150, 210.

1. n=11 ⇒ median Q2 = 6th value = **60**.
2. Position rule (n+1)/4 ⇒ Q1=3rd=**42**; Q3=9th=**90**.
3. IQR=90−42=**48**; 1.5IQR=**72**.
4. Fences: 42−72=**−30**; 90+72=**162**.
5. Only 210>162, so **210 is circled**; 150 remains inside.

ANSWER

```
five-number summary=(28,42,60,90,210)
fences=(-30,162); circled: 210 only
```



Source redraw with the worked values: the whisker ends at 150; 210 is only flagged.

TRAP	CORRECTION
circle = delete	no: investigate context first
whisker = max	only most extreme value inside fence
mean resistant	median/IQR resist extremes; mean/sd do not

14 · EDA fast recall

CH03

MOMENT	MEANING	ROBUST PARTNER
1st	location / mean	median
2nd	scale / variance; report sd	half-IQR
3rd	symmetry / skew	Galton skewness
4th	tails / kurtosis; Normal=3	Moors kurtosis

Mean chases the long tail: right-skew ⇒ mean>median; left-skew ⇒ mean<median. Q-Q points near y=x support Normal residuals; tail departures signal skew/heavy tails.

MINI DRILL	CORRECT CLASSIFICATION
500-listener price survey	small data: bounded, deliberate, once-off
continuous play/skip/search/device logs from tens of millions	big data: Volume + Velocity + Variety
quarterly revenue report	small data
9-person listening lab	small data

Answer structure: state the four-V test; eliminate bounded deliberate sources; select the continuous multi-source logs. Rigour and byte count do not decide the label.

MISSING-DATA CUE	MECHANISM
independent of every variable	MCAR
depends on visible age/group	MAR

15 · Drill 2: vectors

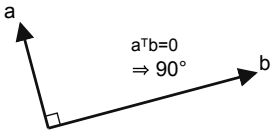
CH04 WORKED EXAMPLE

a=[3,1,−2]ᵀ, b=[2,k,4]ᵀ.

1. ⟨a,b⟩=(3)(2)+(1)k+(−2)(4)=6+k−8=**k−2**.
2. Orthogonal ⇒ inner product zero: k−2=0 ⇒ **k=2**.
3. Then b=[2,2,4]ᵀ and ‖b‖₂=√(2²+2²+4²)=√24.
4. √24=**2√6≈4.90** (3 s.f.).

ANSWER

```
(a,b)=k-2; k=2; ‖b‖₂=2√6≈4.90
```



Source redraw: "orthogonal" is an instruction—set the dot product to zero and solve.

SELF-CHECK	EXPECTED
inner product type	scalar
norm sign	non-negative
‖b‖ vs ‖b‖²	√24 vs 24

16 · Drill 3: K-means

CH05 WORKED EXAMPLE

A=[1,1], B=[2,1], C=[6,5], D=[7,5]; k=2; μ₁=[1,1], μ₂=[2,1].

POINT	D² TO M₁	D² TO M₂	ASSIGN
A	0	1	C₁
B	1	0	C₂
C	41	32	C₂
D	52	41	C₂

C₁={A}; C₂={B,C,D}. μ₁=[1,1]; μ₂=[(2+6+7)/3,(1+5+5)/3]=**[5,3.67]**. New WCSS ≈**24.7**, lower than the start; next pass flips B into C₁.

WRITE: *squared distance; mean of new members; objective non-increasing; local optimum.*

K-MEANS CLAIM	VERDICT
supervised	false: there is no target y
centroid must be observed	false: it is a coordinate-wise mean
chooses k	false: k is input
objective strictly falls	false: non-increasing; may hold
global optimum guaranteed	false: initialisation matters

Squared distance and ordinary distance rank centres identically, but squaring removes the root and makes the mean the clean update minimiser.

16B · Matrix product drill

CH06 WORKED EXAMPLE

GIVEN

```
A=[[2,1],[3,0]], B=[[1,4],[5,2]]
```

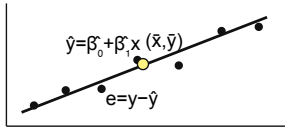
Row·column gives AB=[[2·1+1·5, 2·4+1·2],[3·1+0·5, 3·4+0·2]]=**[[7,10],[3,12]]**.

17 · Drill 4: OLS + R²

CH06 WORKED EXAMPLE

x̄=10, ȳ=50, S_{xx}=200, S_{xy}=900, TSS=5000.

1. $\hat\beta_1 = S_{xy}/S_{xx} = 900/200 = \mathbf{4.5}$.
2. $\hat\beta_0 = \bar y - \hat\beta_1\bar x = 50 - 4.5\times10 = \mathbf{5}$.
3. Fitted line: **ŷ=5+4.5x**.
4. SSR=$\hat\beta_1 S_{xy}$=4.5×900=**4050**.
5. R^2=SSR/TSS=4050/5000=**0.81**; RSS=950.
6. +1 in x is associated with average +4.5 in y; x explains 81% of variation.



Source redraw: the OLS line passes through (x̄,ȳ) and minimises squared vertical residuals.

TRAP	REPAIR
causal wording	say "associated with"
R^2=81 units	81% of variation explained
Normality needed to fit	only needed for inference

18 · Regression diagnosis script

CH06–07

SEE	STATE	THEN
random even band	mean + variance plausible	model adequate
U/curve	E(ε|x)=0 fails	polynomial/basis term
fan/funnel	var(ε|x)=σ² fails	log response
right-skewed y	tail + variance issue	log y; correct back-transform

Answer in three clauses: shape → assumption → remedy. "The model is bad" earns little without the named failure.

SECOND OLS SPEED DRILL

```
x=(1,2,3,4), y=(2,2,4,5)
x̄=2.5; ȳ=3.25; Sₓₓ=5; Sₓᵧ=5.5
β̂₁=1.1; β̂₀=0.5; ŷ=0.5+1.1x
fitted=(1.6,2.7,3.8,4.9)
```

Residuals=(0.4,−0.7,0.2,0.1), so RSS=0.70. The fitted values and residual signs provide a quick arithmetic check.

FINISH THE FIT CHECK

```
TSS=Σ(yᵢ-ȳ)²=6.75
R²=1-0.70/6.75≈0.896
```

OLS QUANTITY	MEANING
$e_i = y_i - \hat y_i$	vertical residual
RSS=Σe_i^2	loss minimised
R^2	fraction of y variation explained in sample
SE(β̂)	sampling variability for inference

TRANSFORM CLUE	FEATURE MOVE
right-skewed positive y	log y, then smear on return

19 · Drill 5: confusion matrix

CH08 WORKED EXAMPLE

y=[1,1,0,1,0,0,1,0,1,0]; ŷ=[1,0,0,1,1,0,1,0,1,1].

	ACTUAL 1	ACTUAL 0	TOTAL
PRED 1	TP=4	FP=2	6
PRED 0	FN=1	TN=3	4
TOTAL	5	5	10

1. Accuracy=(4+3)/10=**0.70**.
2. Recall=4/(4+1)=**0.80**.
3. Precision=4/(4+2)=**0.667**.
4. F1=8/(8+2+1)=8/11=**0.727**.

MARGIN MEMORY

```
recall: actual-column
TP/(TP+FN)
precision: predicted-positive row
TP/(TP+FP)
FPR: actual-negative column FP/(FP+TN)
```

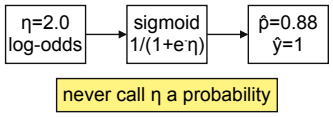
ALWAYS: *draw the 2×2 before calculating. Accuracy can look excellent when the rare positive class is never found.*

20 · Logistic worked prediction

COURSE WORKED EXAMPLE

β_0=−2.0, β_1=2.0, β_2=1.0; x_1=1.5, x_2=1.

1. η=−2+(2)(1.5)+(1)(1)=**2.0** log-odds.
2. p̂=1/[1+exp(−2)]; exp(−2)≈0.135.
3. p̂=1/1.135≈**0.88**.
4. 0.88>τ=0.5 ⇒ **ŷ=1, churn risk**.



Source redraw: linear predictor first, sigmoid second, threshold last.

COEFFICIENT	INTERPRETATION
β_j>0	raises log-odds and p
exp(β_j)	odds multiplier for +1 in x_j
τ lower	TPR↑ and FPR↑

20B · Ethics case drill

CH07 WORKED EXAMPLE

A résumé screen down-ranks certain postcodes, cannot explain rejections, and later shows systematic ethnic-group disadvantage.

PRINCIPLE	ONE-SENTENCE JUSTIFICATION
Outcomes	postcode acts as a proxy and produces disparate impact regardless of legitimate efficiency intent
Transparency	rejected candidates cannot understand or contest how their data drove the decision

21 · Validation MSE

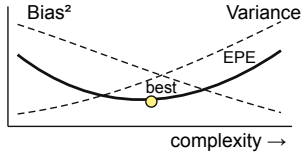
CH09 WORKED EXAMPLE

Validation y=(8,12,20,15); ŷ=(10,11,17,14).

RUN THE ARITHMETIC

```
e=y-ŷ=(-2,1,3,1)
e²=(4,1,9,1); Σe²=15
MSE=(1/4)·15=3.75
```

MSE is the mean, not the sum. Compute it for every candidate; choose the lowest validation MSE. Test MSE is reported once after selection.



Source-grounded redraw: bias falls, variance rises and EPE is minimised in the middle.

REGIME	BIAS	VARIANCE	SIGNAL
under-fit	high	low	train + validation poor
good	balanced	balanced	validation minimum
over-fit	low	high	train low, validation rises

22 · Poisson MLE

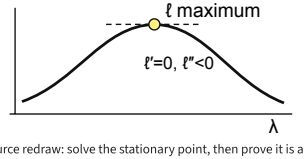
CH10 WORKED EXAMPLE

y_i iid Poisson(λ); sample (2,4,3,5,1), n=5, Σy_i=15.

FIVE MARK LINES

```
L(λ)=∏ᵢ λ^yᵢ e^−λ/yᵢ!
ℓ(λ)=(Σyᵢ)logλ-nλ-Σlog(yᵢ!)
dℓ/dλ=(Σyᵢ)/λ-n=0
λ̂=(Σyᵢ)/n=ȳ=15/5=3
d²ℓ/dλ²=-(Σyᵢ)/λ²<0 ⇒ maximum
```

Same skeleton: Bernoulli MLE = sample proportion; exponential rate MLE = 1/ȳ. OLS equals MLE under Gaussian errors.



Source redraw: solve the stationary point, then prove it is a maximum.

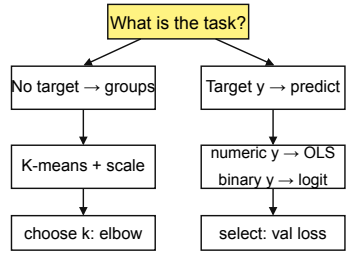
WHY LOG L?	PAY-OFF
many probabilities <1	avoid underflow
product	becomes differentiable sum
log increasing	same argmax

TALL-DATA EXTENSION: N=2,000,000

```
full batch reads all rows ⇒ O(n)/step
SGD uses one random row ⇒ O(1)/step in n
up to n cheap updates per full-pass budget
```

23 · Method chooser

CH03–11 SYNTHESIS



Start from the response and task, not from a favourite algorithm.

QUESTION WORDING	IMMEDIATE MOVE
"orthogonal"	x^Ty=0; solve
"nearest centroid"	compare squared distances
"best line"	minimise RSS / OLS
"probability of class 1"	η then sigmoid
"rare positive"	precision / recall / F1
"unseen performance"	validation loss; reserve test
"most plausible parameter"	likelihood → log → differentiate
"millions of rows"	tall; SGD / row blocks
"p>n"	wide; cut predictors

24 · Train / select / report

CH09

1 · TRAIN	2 · VALIDATE	3 · REFIT	4 · TEST
fit all candidates	choose lowest loss	chosen form on train+val	report once

The test set is not a second validation set. Repeated use leaks information and makes the reported performance optimistic.

25 · Python from memory

CH04, 08, 09

NUMPY ESSENTIALS

```
import numpy as np
x=np.array([3,1,-2])
np.dot(x,y)   # inner product
mse=np.mean((y-yhat)**2)
```

SKLEARN SKELETON

```
model.fit(X_train,y_train)
p=model.predict_proba(X_val)[:,1]
yhat=(p>0.5).astype(int)
```

Vectorise rather than loop. Check brackets, axis, array shapes and imports. Code marks reward runnable syntax, not prose.

26 · Final trap scan

ALL CHAPTERS

- AB≠BA; check inner dimensions.
- K-means: squared distance; local, not guaranteed global.
- Regression slope: association, not cause.
- η=$x^T\beta$ is log-odds, not probability.
- Training error is not a selection criterion.