

This report summarises the types of questions in the sample exam and past papers from 2020 to 2021 (not including mastery questions). Each number in the bracket indicates the number of marks this type of question appeared.

Chapter	Questions	Total Marks
Block I: Regression and Variable Selection	<ul style="list-style-type: none"> - Definition (2) <ul style="list-style-type: none"> - unbiased estimator - Gauss-Markov theorem - Prove theorems in lecture notes: (3) <ul style="list-style-type: none"> - Gauss-Markov theorem - Linear Regression (28) <ul style="list-style-type: none"> - write optimisation problems: least square, ridge, lasso - derive estimators: OLS, ridge, LASSO - inference on estimators: mean, variance, confidence interval - Explanation for estimators <ul style="list-style-type: none"> - Motivations to define, characteristics - Bayesian formulation of ridge regression - Practical Question: Interpretation of R code output (7) <ul style="list-style-type: none"> - linear model summary, plot of regression estimator against regulator, correlation matrix 	40 20%
Block II: Distance- based Methods	<ul style="list-style-type: none"> - Multidimensional Scaling (23) <ul style="list-style-type: none"> - inner product matrix, distance matrix, configurations etc. - loss of information - Eigenvalues - Apply Miles algorithm (2) - Analyse clustering (4) - Explanation (8) <ul style="list-style-type: none"> - Procedures/Purpose of models/methods - Practical Question: Interpretation of R code output (3) <ul style="list-style-type: none"> - Eigenvalues, recovered configuration 	40 20%
Block III: Modern Regression	<ul style="list-style-type: none"> - Book work <ul style="list-style-type: none"> - examples of basis expansions (1) - Kernel density estimator (2) - Problem rewriting (6) <ul style="list-style-type: none"> - Rewrite RSS, rewrite smoothing spline as ridge regression - Wavelets (24) <ul style="list-style-type: none"> - Derive orthogonality, estimator for coefficients, independence between noises, computational efficiency, interpretations of parameters, posterior distribution - Kernel DE (18) <ul style="list-style-type: none"> - Inference of estimator: Expectation, Variance - derive estimator - Explanation (6) <ul style="list-style-type: none"> - Meaning of model, regulator, Choice of model - Practical: interpret results (5) <ul style="list-style-type: none"> - mean squared error, residual plots, fitness plot 	64 32%
Block IV: Unsupervised Learning	<ul style="list-style-type: none"> - Definition (2) <ul style="list-style-type: none"> - Kullback-Leibler divergence, negative entropy - Derive relationship between normal distribution and negative entropy (5) - Compare two statistical models (4) <ul style="list-style-type: none"> - Projection Pursuit vs Independent component analysis - Centralisation, spherling of data (5) 	16 8%

Chapter	Questions	Total Marks
Block V: Trees, Bootstrap, Bagging	<ul style="list-style-type: none"> - Definition (3) <ul style="list-style-type: none"> - Training/test sets - AdaBoost.M1 - Classification/regression tree (21) <ul style="list-style-type: none"> - Derive estimator(minimiser of loss), best split point, why bi-partition, split criterion and minimisation of it, reasons to use classification trees, example where classification tree fails - AdaBoost (6) <ul style="list-style-type: none"> - find minimiser, explain choice of operator - Practical Question: interpret output of R function (4) <ul style="list-style-type: none"> - tree (R function) 	34 17%
Others	<ul style="list-style-type: none"> - Derive an equation/inequality (4) - sketch function (2) 	6 3%

Distribution of Marks by Question Type

Bookwork(writing down definitions, state theorems from lecture notes):	21 (10.5%)
Calculation, Derivations	93 (46.5%)
Mathematical Proof (like analysis, linear algebra)	12 (6.0%)
Explanation, interpretation	72 (36.0%)
– of which: interpret R output	19 (26.4%)
Others	2 (1.0%)

Top Questions

- Linear Regression
- Wavelets
- Multidimensional Scaling
- Classification/regression trees
- Kernel DE

List of Models/Algorithms:

- **Block I** Ordinary Least Square (Variants: ML, ML Gaussian, Heavy tails) Multivariate Linear Model, Forward selection, backward deletion, Ridge Regression, LASSO, Principal Components Regression
- **Block II** Classical Scaling, k-mean clustering, self-organising maps, Procrustes Analysis
- **Block III** Piecewise Polynomial, Smoothing splines, Kernel Smoothing, Kernel Density Estimation, Nadaraya Watson Estimator, Local Polynomial Regression, Orthogonal Series Expansion, Wavelets (Discrete Wavelet Transform, Wavelet Shrinkage, Bayesian Wavelet Shrinkage), PCA, Projection Pursuit, Independent Components Analysis
- **Block IV** Projection Pursuit Regression, Neural Networks
- **Block V** Regression Tree, Classification Tree, Bagging Trees, Random Forests, Boosting, AdaBoost