TOPIC			Notebook							
			linear regression	**JN1**	**JN2**	**JN3**	**JN4**	**JN5**	**ML4HAR_1**	**ML4HAR_2**
			Univariate regression linear dataset with noise --> linear regression quadratic dataset with noise --> polynomial regression extension to multivariate regression: >1 input features Evaluating model performance: overfitting	linearly separable datasets binary classification Binary classification methods with 2 features for easy visualisation evaluating model performance	non-linearly separable synthetic datasets Binary classification methods with 2 features for easy visualisation Evaluating model performance:	binary classification effects of class imbalance undersampling, over-sampling	a more realistic dataset: predicting academic performance Pandas dataframes to handle the training set multi-class problem > 2 features Overfitting Cross-validation Ensemble methods to mitigate overfitting in decision trees: Random Forests	Classification methods for > 2 classes and > 2 features Visualising higher dimensional data: PCA Controlling overfitting - repairing class unbalance - regularisation and penalties - class weights Model selection and optimisation - tuning hyper-parameters using cross validation	HAR on PAMAP dataset with high level features and ad hoc feature engineering. small dataset	HAR on PAMAP dataset with high level features and ad hoc feature engineering. FULL dataset
synthetic dataset generation			x							
	linear data		x							
	polynomial data		x							
	linearly separable			x		x				
	non linearly separable				x					
	unbalanced classes					x				
real datasets										
	academic performance									
dataset visualisation				x	x	x				
	using PCA							x		
pandas dataframes							x			
problem type										
	regression									
		univariate	x							
		multivariate	x							
	classification									
		logistic regression		x	x	x	x			
		SVM (linear)		x	x	x	x			
		decision trees		x	x	x	x			
		SVM (polynomial)			x		x			
		SVM (RBF)			x		x			
		random forests					x			
problem shape										
	outcome									
		binary	x	x	x	x				
		multi-class					x			
	features									
		2 features		x	x					
		many features					x			
pre-processing										
	feature engineering: linear --> polynomial		x							
	scaling			x	x					
	downsampling					x				
	upsampling (SMOTE)					x	x			
	one-hot encoding						x			
univariate dataset visualisation			x	x						
model evaluation			x							
	MSE, R^2		x							
	accuracy			x	x		x			
	F1			x	x		x			
	AUC / ROC curves			x	x		x			
	average precision score			x	x		x			
training / test split			x	x	x		x			
	cross validation for performance									
overfitting / generalisation error			x							
	regularisation / ridge		x							
	regularisation / lasso		x							
	regularisation / ElasticNet		x							
hyper-parameter tuning							x			
	gridSearch with cross-validation for optimisation						x			