

Argomento	Sottoargomento	Sigla	Domande
LM e GLM: Logistic Regression & LDA, QDA	R ²	SM	Che cosa accade a R ² quando vengono aggiunte al modello lineare covariate irrilevanti.
		SM	cos'è l'adjusted R ² (pt ignoti)
		M	Descrivere un metodo per valutare la bontà di un modello di regressione
		LM	A cosa serve l'analisi dei residui e strumenti per la sua effettuazione
		LM	Come determinare la qualità di una previsione in un modello lineare (non ricordo bene il testo della domanda, né se c'erano ulteriori richieste)
	LR	M	Cosa si intende per codifica dummy dei predittori categoriali nel modello di regressione
		LM	interpretazione dei coefficienti nella multivariate Linear regression (pt ignoti)
			Perché si chiama regressione lineare?
		SM	Perché la regressione lineare non è adatta alla classificazione
			Cosa è la Normality assumption?
	Log Reg	LM	Cosa rappresentano i coefficienti nella regressione logistica lineare.
		SM	How do you use fitted logistic regression for prediction?
		LM	Confronto tra LDA e Logistic Regression
	DA	M	Discriminant analysis
		SM	Che cosa significa il "Linear" in LDA
	LDA, QDA	LM	Confronto tra LDA e QDA
			Cosa è un Bayesian Classifier?
Principal Component Analysis	PCA	L	Descrivere come computare le componenti principali
	PCA	LM	Descrivi un metodo per calcolare le componenti principali
	PCA	LM	Come calcolare le PCA
		SM	Perché è importante la dimension reduction? Fare una lista dei motivi
Clustering		L	Descrivi metodi alternativi per misurare la distanza del gruppo e possibili effetti sul risultato del cluster
		M	Si descrivano metodi alternativi per misurare la distanza tra gruppi e i possibili effetti sui metodi della Cluster Analysis
		SM	A method for calculating k cluster in non-hierarchical method?
		SM	Descrivere un metodo non gerarchico per la cluster analysis
		LM	Descrivi metodi alternativi per misurare la distanza fra i gruppi e i possibili effetti sui risultati del cluster
	K-means	SM	Cos'è la matrice delle distanze?
		SM	Modo di inizializzazione di K-means
		SM	descrivi un criterio per selezionare il numero di cluster in un clustering non gerarchico.
	KNN		Perché KNN non funziona per P grandi?
		SM	Modo per determinare il numero K di cluster nel clustering non gerarchico
	FP, FN		Definire e descrivere tre indici per la classificazione binaria
			Cosa sono FP, FN e come sono in relazione?
	Decision Tree	LE	Obbiettivi e metodologia della cross-fold validation
		L	Spiegare obiettivi e metodologia della cross fold validation
		L	Descrivere un criterio di arresto appropriato per l'apprendimento dell'albero decisionale. Descriverne l'impatto
		X	Describes a proper stopping criterion for learning decision tree.
		S	Describes impact on
			What is node inpurity in classification/regression problems

	Decision\	S	Cosa si intende per impurità di un nodo in un albero di regressione/classificazione?
	Regression tree	S	Impurità di un nodo nel caso di regressione classif
	Regression tree	S	Criterio di stop per la generazione albero di regressione
		E	Cosa succede nel Random Forest se si aggiungono variabili esplicative poco rilevanti?
	RF	S	M in RF? qual è un buon valore?
		SM	Che cos'è uno strong predictor?
	Bagging	E	Bagging. Gli alberi devono essere potati?
		SE	Caso Bagging trees gli alberi devono essere potati? Perché?
	RF, Bagging	LE	Nel bagging bisogna potare ogni albero?
		E	Nel bagging bisogna sempre potare? Perché? (pt ignoti)
		S	What is OOB Error?
	Accuracy	L	Describes and plot a visual method for riconoscere l'overfitting in un modello di classificazione supervisionata
		E	Accuracy di un classificatore binary sue pecche
		LM	Discutere (con un esempio) le limitazioni dell'accuratezza come indice di valutazione di sistemi di classificazione binaria
		1	Legame tra FPR FNR e Accuracy
		S	What is Accuracy@K?
EER	L	A cosa serve l'analisi dei residui e strumenti per la sua effettuazione	
	S	Equal error rate	
		Descrivere un metodo di calcolo dell'indice AUC	
AUC	SE	AUC	
	L	Come calcolare AUC in classificazione binaria	
	LE	Dire un metodo con cui calcolare AUC	
SVM	Multiclasse	LE	Descrivi come si può effettuare una classificazione multiclasse con SVM
		LE	Descrivi come si può effettuare una classificazione multiclasse con SVM
	Soft margin	SE	Si può apprendere un soft margin classifier per ogni possibile problema di classificazione binaria? Perché?
		SE	Con il soft margin classifier si può sempre fare classificazione binaria?
	Variabili categoriche	S	SVM può lavorare con variabili indipendenti categoriali?
		S	SVM non si possono utilizzare con variabili dipendenti categoriali
	Maximal Margin classifier		Il SVM si possono usare con variabili risposta categoriche?
		SE	il maximal margin classifier è adatto a tutti i problemi di classificazione binaria?
Text Mining		SE	Maximal Margin Classifier può essere usato per tutte le classificazioni binarie? Why? (domanda da controllare)
		ES	Is stemming language independent? Why?
		S	Lo stemming è indipendente dalla lingua (corta)
		L	Descrivere almeno 3 modi per fare preprocessing in text mining
		LE	si descrivano almeno 3 operazioni di pre-processamento di dati t
		S	La sentiment analysis è regressione? perché?
		S	Si può porre la sentiment analysis come un problema di regressione? Come?
			Classificare per argomento dei temi scolastici:hai un problema di classificazione di temi per argomento e devi proporre un modo per risolverlo (Text mining)
			No free lunch (spiegazione della domanda è quel teorema secondo cui non esiste un classificatore che risolve qualsiasi problema in modo ottimale. Si rifà a quel discorso sui menù e i ristoranti.
		LE	Si discuta il teorema del No free Lunch nello scenario degli algoritmi di apprendimento

		S	Si può porre la sentiment analysis come un problema di regressione? Come?
Topic Modelling	Corpus	L	Dato un corpus di 1.000.000 di documenti in inglese, italiano e un'altra lingua, descrivere un metodo per stabilire i temi più rilevanti
	LDA	S	Is Latent Dirichlet Allocation supervised?
		S	Is Topic Modelling a supervised approach? Why?
		S	il topic modeling è un approccio supervisionato?
Recommender System	serendipity	S	Is serendipity a performance metric suitable in binary classification? Why?
		L	Spiegare almeno tre metodi di assessment dei sistemi di classificazione binaria
	Collaborative filtering	L	Problem statement del collaborative filtering
		LM	Che problema risolve il collaborative filtering?
	Diversity	S	Is diversity a performance metric suitable in univariata regression? Why?
Evolutionary Computation	Pareto	L	Cos'è la dominanza di Pareto?
		LE	Spiegare e fare un esempio dell'ottimo di Pareto
		L	Cos'è la dominanza di Pareto ?
		L	Spiegare e fare un esempio della dominanza di Pareto
	GP (Genetic Programming)	S	Come si rappresenta il fenotipo nei GP?
		SE	Come viene rappresentato il fenotipo in GP?
		S	Cos'è genotipo in GP
		S	Cosa è il fenotipo nei GA
	Genotipo e fenotipo	SE	Qual'è la rappresentazione fenotipica
		d	Differenza tra genotipo e fenotipo ed esempi
			Descrivi il rapporto tra Genotipo e Fenotipo.
	Selection criteria	LE	Come funziona la tournament selection?
		S	Descrivere almeno un selection criteria
	Exploitation/Exploration	L	Descrivere il trade-off per exploration ed exploitation nell'evolutionary computation
		S	Exploitation/Exploration in EC
	Population dynamics	S	Mention one typical generational model in EC
			Citare un generational model
	Crossover		Crossover
		S	Nominare 1 crossover operator e disegnare un esempio
	GE (Grammatical Evolution)	S	Phenotype in GE?
		S	Method for estimation of training error
Goodness\Quality of fit	test\ training		Descrivi cosa sono test e training sets
		manca	Si fornisca un metodo per stimare l'errore di training di classificazione
		S	Cos è il training error? Fare un esempio
	Model Performance	LM	Descrivi come è possibile stimare la performance di un test-set tramite metodi indiretti. Elenca i criteri di selezione alternativi.
		L	Describes a proper procedure to evaluate a model predictive performance
			Come valutare la capacità predittiva di un modello
		S	Modo per valutare bontà di un modello predittivo
		S	The principal aims of classification methods

	L	descrivi le principali finalità di un modello di classificazione
		Cos'è l'overfitting (dal punto di vista grafico)?
	L	Describes and plot a visual method for riconoscere l'overfitting in un modello di classificazione supervisionata
	LM	Come valutare il modello di apprendimento: si descriva una procedura Descrivere un modo (anche graficamente) per vedere overfitting