QBUS3820 Mid-Term Sample

	Question:	Answer A:	Answer B:	Answer C:	Answer D:	Answer E:
1	In traditional statistics, a parameter is	a random quantity, a function of the sample	a fixed value to be estimated	is the average height of Australians		
2	A regression is an example of	unsupervised learning	supervised learning	data mining	A and C	B and C
3	Let a'=[1, -1, 2] and b'=[2, 1, -2]. What is the inner product a'b and what is the norm of a?	-3 and 6	-1 and 6	-3 and square root of 6	-1 and square root of 6	-3 and 3
4	What is the Euclidean distance between a'=[5, 3, -2] and b'=[3, 1, -1]?	23	11	3	sq root of 11	sq root of 20
5	What does this formula represent and how would you write it in terms of inner products and vector norms of column vectors x and y? $$\frac{\sum_i x_i y_i}{\sqrt{(\sum_i x_i^2)(\sum_i y_i^2)}}$$	sample covariance and x'y/||x||^2	sample correlation and xy/sqrt(||x|| ||y||)	sample correlation and x'y/(||x|| ||y||)	none of the above	
6	Let X'X = [1, 1; 1, 1.01]. What is the inverse of X'X?	[1.01, -1; -1, 1]	[101, -100; -100, 100]	[100, -100; -100, 101]		
7	Which of the following statements are true for any matrices A: NxP, B:NxN, C:PxN	A'B=B'A	BAC = CBA	tr(BAC) = tr(CBA)	1 and 3	none of the above
8	A training data set is a sample on which we test a learner	no	yes			
9	k-NN estimator produces no mis-classifications if k equals	1	infinity	some positive number not equal to 1		

10	Why do we want to have at least two samples from the same DGP?	to check if a model generalizes	to avoid overfitting, i.e. fitting a model with more parameters than needed	because if we have one sample the training error will always go down if we add parameters but that is not true for a new sample.	all of the above	
11	Conditional expectation or mean function of Y given X is	the expected value, or population average, of Y when X takes a specific value	the sample average of Y for values of X, in a neighborhood of a specific value	a weighted average of sample values of Y	none of the above	
12	OLS produces an estimate of the conditional expectation of Y given X	replacing averages over the population with averages over a sample of Y and X	assuming the conditional expectation function is linear in X	that has potentially a high bias if linearity fails but is more precise due that assumption	that is consistent asymptotically if the linear model is correctly specified	all of the above
13	Maximum Likelihood Estimator	finds the value of parameters for which the likelihood of those parameters is highest	finds the value of parameters for which the likelihood of observing the sample is highest	is equivalent to OLS if based on the assumption of normal errors	A and C	B and C
14	Penalty terms in nonparametric estimators are usually used to	restrict the number of parameters	reduce variance of the estimator	induce bias	B and C	all of the above
15	There is a way to find optimal model complexity at which	training error is minimal and this can be done by using a test set or an information criterion	testing error is minimal and this can be done by using a test set on the same model or using an information criterion	testing error is minimal and this can be done by using a resampling technique (such as bootstrap or cross-validation) or by using an information criterion	A and B	B and C
16	AIC penalizes large models more than BIC	TRUE	FALSE			
17	What alternative methods can be used for classification, aside from SVM	Linear Discriminant Analysis	Quadratic Discriminant Analysis	Logistic Regression and Probit	A and B	All of the above

18	What does the Bayesian rule say?	That class posterior is proportional to density of regressors given their class	That the probability of belonging to class j given X is equal to density of X given it is in class j, times the probability of class j, over the density of X	That the probability of belonging to class j given X is equal to the density of X given it is in class j, times the probability of class j, divided by the sum of products of densities of X for each class, and the probabilities of each class	A and C	All of the above
19	In discriminant analysis, the discriminant function for class k is	log of the posterior probability of class k	log of the product of the prior probability of class k and density of X belonging to class k	depends on the assumed normal distribution for the density of X	A and C	All of the above
20	How would you estimate the prior probabilities of belonging to a class?	As a proportion of units belonging to each class	As a result of Bayesian updating from posterior probabilities	We can take any prior probabilities of belonging to a class, they don't matter to LDA, QDA or anything else.		
21	The empirical error rate for the training sample of size N is	the ratio of false positives and false negatives to N	the ratio of true positives and true negatives to N	the ratio of misclassified observations to N	A and B	A and C
22	Suppose the odds ratio of getting cancer is 3 (ie "3 to 1 "). What is the probability of getting cancer?	0.333333333	0.75	0.66	0.33	Both A & C