## Quiz 9

Started: Nov 29 at 6:15pm

## **Quiz Instructions**

This quiz is open book, open notes, "open R". The expected duration is 60 minutes. Two attempts are allowed. If both attempts are taken, the score for the second attempt will "overwrite" that from the first attempt (regardless if it is higher or lower). Even though the quiz has 24 points, it will be graded out of 20 points (i.e., 4 points bonus).

You are allowed to use any of the class materials from our SML class, but no other materials (no internet browsing or communication with other parties online/offline).

Even if a question is asking for a numerical value or True/False answer, in order to receive full credit (if your "guess" is correct) or partial credit (if appropriate, if your "guess" is incorrect), please provide your rationale as comments in the uploaded file requested at the end of the guiz.

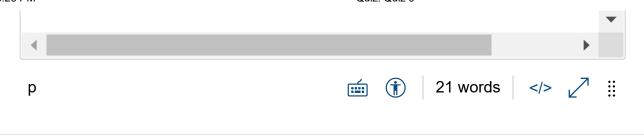
Question 1 4 pts

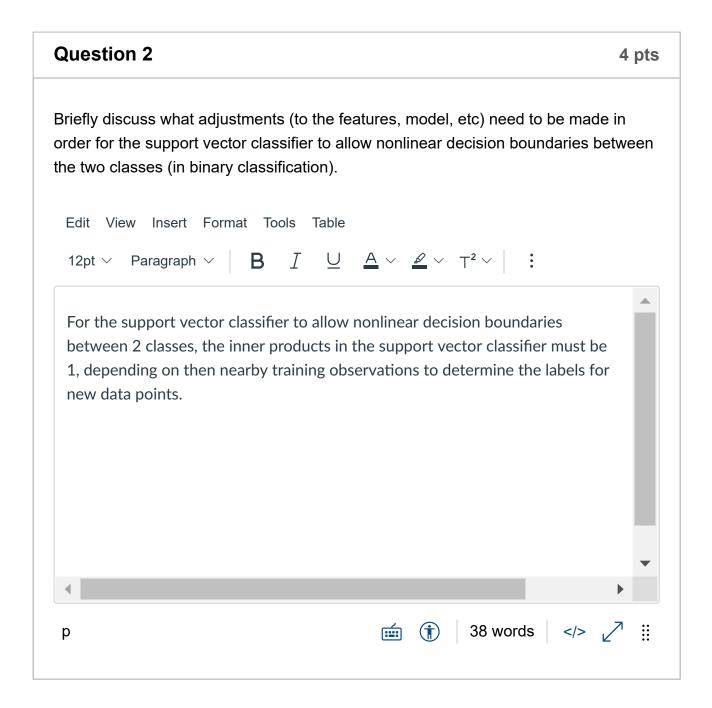
Consider a binary classification problem with two covariates, x1 and x2, using a maximum margin classifier with the separating hyperplane f(x)=0, where  $f(x)=x_1-2x_2+1$ . Will the following two new data points with features  $(x_1,x_2)=(1,2)$  and  $(x_1,x_2)=(0,0)$  be assigned to the same class or to different classes? Briefly explain.

Edit View Insert Format Tools Table

12pt  $\vee$  Paragraph  $\vee$  B I  $\underline{\cup}$   $\underline{\wedge}$   $\vee$   $\underline{\mathscr{A}}$   $\vee$   $\top^2$   $\vee$  :

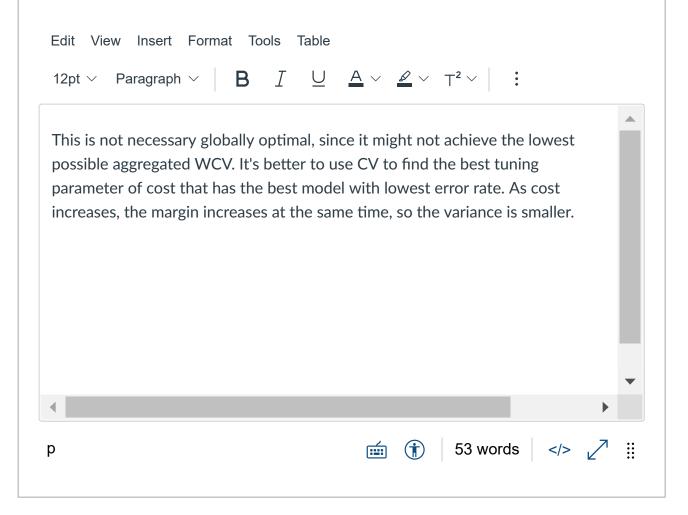
f(1,2)<0, f(0,0)>0, thus using this seperating hyperplane, these 2 data points will be assigned to different classes.

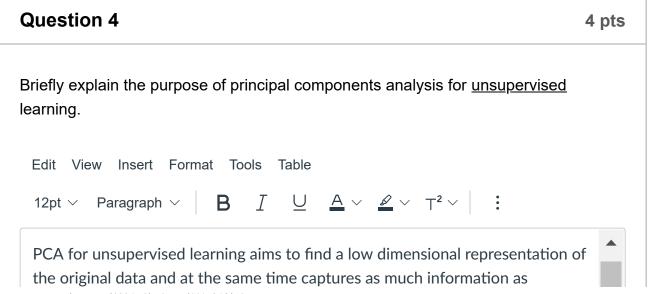


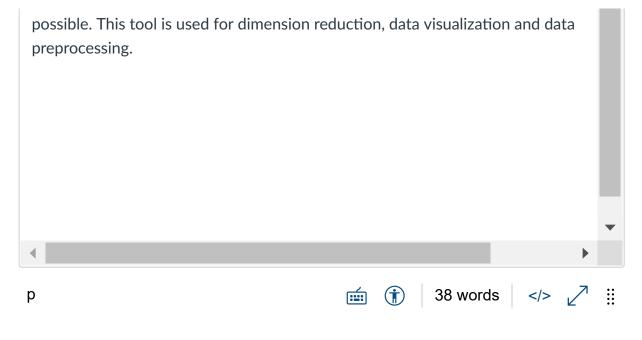


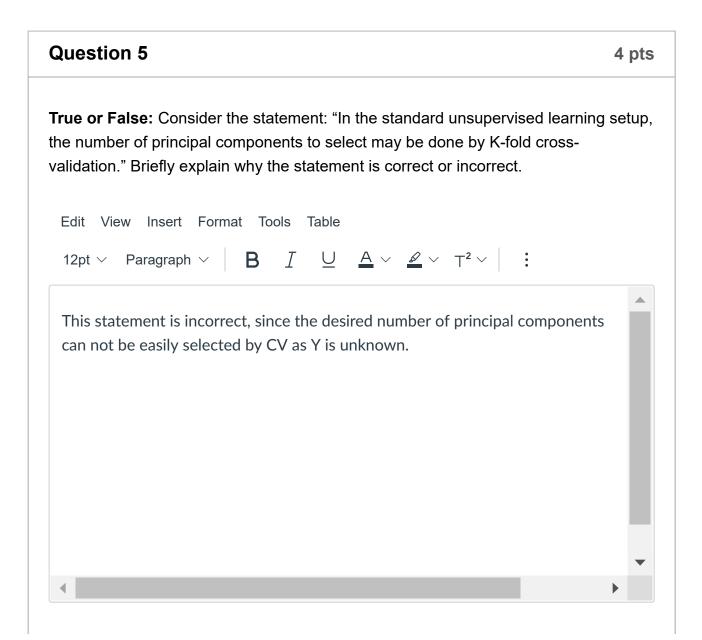
Question 3 4 pts

Suppose the K-means algorithm has been run <u>once</u> till convergence to assign data points to, say, K=3 clusters. Is this solution (globally) optimal (for the same value of K), i.e., achieves the lowest possible aggregated within-cluster variation (WCV) for these data? What additional information/experiments would be helpful in answering this question?

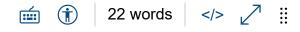








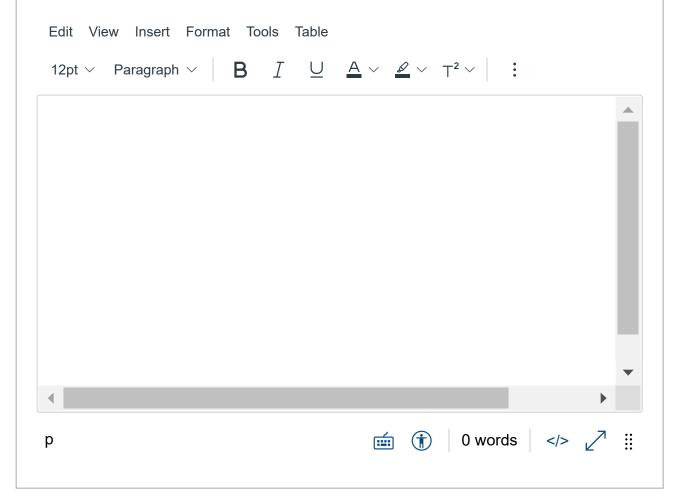
p



Question 6 4 pts

Let X be the design matrix with p numerical features (columns) and n observations (rows). Assume that n > p > 2. Also assume that the columns of X have been centered (to have mean zero) and rescaled (to have variance 1). Refer to the rows of X as "observations".

Let  $X = U \cdot D \cdot W$  be the singular value decomposition of X. Briefly explain how to obtain the first two principal component scores (i) for our observations (data matrix X), as well as (ii) a new observation (xnew; assume that it is a row vector that has been preprocessed in the same way as the data matrix X). Feel free to answer using raw R syntax; i.e., you can access and operate on matrices X, U, D, W but not to call princomp() or any other function to perform PCA.

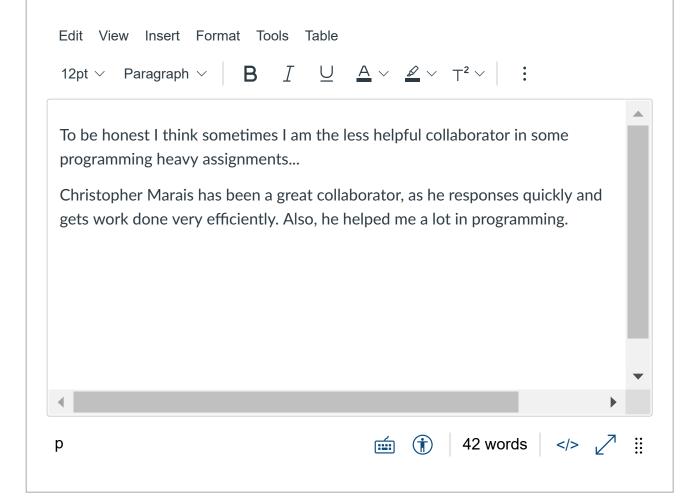


Question 7 0 pts

This question is entirely optional. All student answers are completely anonymous.

Most homework assignments have been collaborative, and based on the working group designs, each student has had a chance to collaborate with every other student in class at least once.

This question is a place to recognize your most (and, if you like, also the least) helpful collaborator during the semester; any relevant details will be appreciated. There are no penalties to the least helpful collaborator(s), but the most helpful collaborator(s) will get a small "karma boost", in case I am undecided about his/her final course grade (if their combined score falls right on the "decision boundary").



Quiz saved at 6:26pm

Submit Quiz