

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 quiz.

Question 1

4 pts

Consider a binary classification problem with two covariates, x_1 and x_2 , using a maximum margin classifier with the separating hyperplane $f(\mathbf{x}) = 0$, where $f(\mathbf{x}) = x_1 - 2x_2 + 1$. 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 ▾ Paragraph ▾ | **B** *I* U A ▾  ▾ τ^2 ▾ | ⋮

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

p



21 words



Question 2

4 pts

Briefly discuss what adjustments (to the features, model, etc) need to be made in order for the support vector classifier to allow nonlinear decision boundaries between the two classes (in binary classification).

Edit View Insert Format Tools Table

12pt ▾

Paragraph ▾

B*I*UA ▾ ▾ T^2 ▾

For the support vector classifier to allow nonlinear decision boundaries between 2 classes, the inner products in the support vector classifier must be 1, depending on then nearby training observations to determine the labels for new data points.

p



38 words



Question 3

4 pts

Suppose the K-means algorithm has been run once 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?

Edit View Insert Format Tools Table

12pt ▾ Paragraph ▾ | **B** *I* U A ▾  ▾ τ^2 ▾ | ⋮

This is not necessary globally optimal, since it might not achieve the lowest possible aggregated WCV. It's better to use CV to find the best tuning parameter of cost that has the best model with lowest error rate. As cost increases, the margin increases at the same time, so the variance is smaller.

p



53 words



Question 4

4 pts

Briefly explain the purpose of principal components analysis for unsupervised learning.

Edit View Insert Format Tools Table

12pt ▾ Paragraph ▾ | **B** *I* U A ▾  ▾ τ^2 ▾ | ⋮

PCA for unsupervised learning aims to find a low dimensional representation of the original data and at the same time captures as much information as

possible. This tool is used for dimension reduction, data visualization and data preprocessing.

p



38 words



Question 5

4 pts

True or False: Consider the statement: "In the standard unsupervised learning setup, the number of principal components to select may be done by K-fold cross-validation." Briefly explain why the statement is correct or incorrect.

Edit View Insert Format Tools Table

12pt ▾ Paragraph ▾ | **B** *I* U A ▾  ▾ τ^2 ▾ | ⋮

This statement is incorrect, since the desired number of principal components can not be easily selected by CV as Y is unknown.

p



22 words

**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 (x_{new} ; 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.

Edit View Insert Format Tools Table

12pt ▾

Paragraph ▾

B*I*UA ▾ τ^2 ▾

p



0 words



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").

Edit View Insert Format Tools Table

12pt ▾ Paragraph ▾ | **B** *I* U A ▾  ▾ τ^2 ▾ | ⋮

To be honest I think sometimes I am the less helpful collaborator in some programming heavy assignments...

Christopher Marais has been a great collaborator, as he responses quickly and gets work done very efficiently. Also, he helped me a lot in programming.

p



42 words



Quiz saved at 6:26pm

Submit Quiz