INFO251 – Applied Machine Learning

Lab 12

Emily Aiken

Announcements

- **PS7** due Monday May 2
- **Quiz 2** on Thursday, April 28
  - **Let us know via email or piazza if you have a DSP accommodation or time conflict**
- **AML hangout –** See Piazza to sign up

Agenda

- Course evaluations
- 5 more practice quiz questions
- Open time for questions

Course Evaluation

- Please fill out the course evaluation if you have not already!
- Our current course evaluation response rate is 20%.
- Visit https://course-evaluations.berkeley.edu.

Practice Quiz Question 11

Regression discontinuity

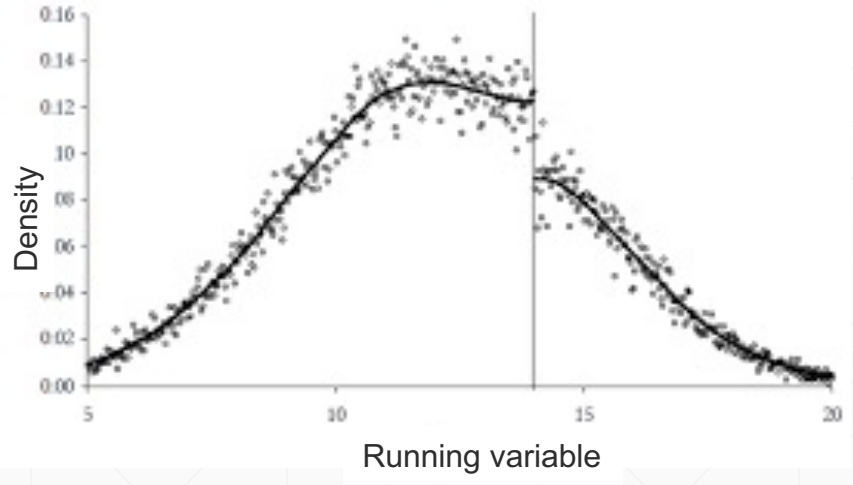
The plot at right of the density of the running variable around a threshold could indicate what for a regression discontinuity design to impact evaluation?

(A) There is visual evidence that the treatment had an impact on the outcome variable.

(B) There is visual evidence that the treatment did not have an impact on the outcome variable.

(C) There is visual evidence that the treatment had an impact on a non-outcome covariate.

(D) There is visual evidence of pre-treatment manipulation of the decision threshold.



Practice Quiz Question 11

Regression discontinuity

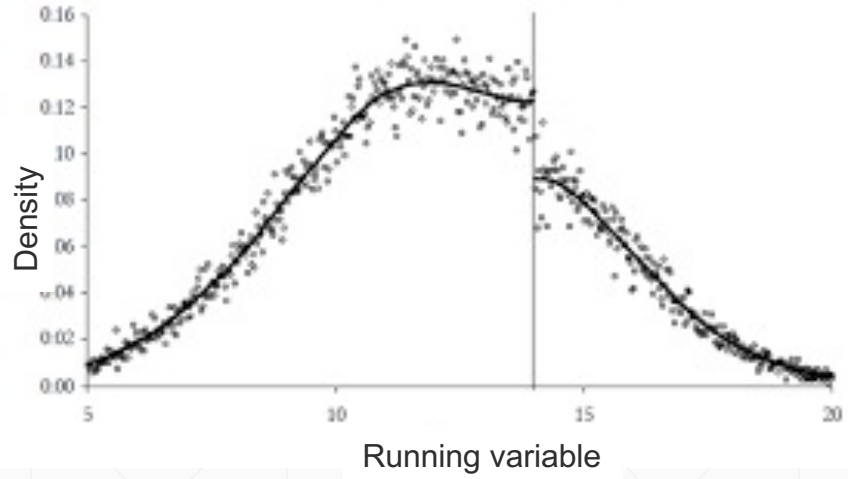
The plot at right of the density of the running variable around a threshold could indicate what for a regression discontinuity design to impact evaluation?

(A) There is visual evidence that the treatment had an impact on the outcome variable.

(B) There is visual evidence that the treatment did not have an impact on the outcome variable.

(C) There is visual evidence that the treatment had an impact on a non-outcome covariate.

(D) There is visual evidence of pre-treatment manipulation of the decision threshold.



Practice Quiz Question 12

Multiclass classification

You are evaluating a classification model for predicting the number in a handwritten digit image from the MNIST dataset. You study examples where the real digit was a 7 but the classifier predicted a 3. This is an example of…

(A) Ablative analysis

(B) Error analysis

(C) Feature importances

(D) SHAP values

Practice Quiz Question 12

Multiclass classification

You are evaluating a classification model for predicting the number in a handwritten digit image from the MNIST dataset. You study examples where the real digit was a 7 but the classifier predicted a 3. This is an example of…

(A) Ablative analysis

(B) Error analysis

(C) Feature importances

(D) SHAP values

Practice Quiz Question 13

Imputation

You are analyzing panel data that tracks poverty over time. You notice that two covariates associated with poverty – education and race – are missing for over 60% of observations in one year of your data. Which would be an appropriate way to deal with the missing data? Select all that would be appropriate.

(A) Drop the observations with missing data

(B) Drop the features with missing data

(C) Model-based imputation, using other covariates to predict education and race

(D) Carry forward education and race from a previous year

(E) Mean, median, or mode imputation of education and race

(F) Zero imputation of education and race

Practice Quiz Question 13

Imputation

You are analyzing panel data that tracks poverty over time. You notice that two covariates associated with poverty – education and race – are missing for over 60% of observations in one year of your data. Which would be an appropriate way to deal with the missing data? Select all that would be at least somewhat appropriate.

(A) Drop the observations with missing data

(B) Drop the features with missing data

(C) Model-based imputation, using other covariates to predict education and race

(D) Carry forward education and race from a previous year

(E) Mean, median, or mode imputation

(F) Zero imputation

Practice Quiz Question 14

Principal components analysis

Which of the following are true about principal components analysis (PCA)? Select all that apply.

(A) The principal components are the eigenvectors of the data's correlation matrix.

(B) The eigenvalues tell you how much variation in the original dataset is explained by each principal component.

(C) PCA is deterministic: If run twice on the same dataset for the same number of components *k,* the results will be the same.

(D) The first PCA component for a decomposition with 1 component will be the same as the first PCA component for a decomposition with 10 components.

(E) PCA should be calculated on standardized features.

Practice Quiz Question 14

Principal components analysis

Which of the following are true about principal components analysis (PCA)? Select all that apply.

(A) The principal components are the eigenvectors of the data's correlation matrix.

(B) The eigenvalues tell you how much variation in the original dataset is explained by each principal component.

(C) PCA is deterministic: If run twice on the same dataset for the same number of components *k,* the results will be the same.

(D) The first PCA component for a decomposition with 1 component will be the same as the first PCA component for a decomposition with 10 components.

(E) PCA should be calculated on standardized features.

Practice Quiz Question 15

Feature importances

Which of the following are methods for calculating feature importances in decision trees, random forests, and other tree-based models? Select all that apply.

(A) Calculate the mean weighted decrease in impurity from splitting on a feature

(B) Count the number of times that a feature is split on in the tree or forest

(C) Permutation importance

(D) SHAP partial dependence plots

Practice Quiz Question 15

Feature importances

Which of the following are methods for calculating feature importances in decision trees, random forests, and other tree-based models? Select all that apply.

(A) Calculate the mean weighted decrease in impurity from splitting on a feature

(B) Count the number of times that a feature is split on in the tree or forest

(C) Permutation importance

(D) SHAP partial dependence plots

Open time for questions