Visualizing Data for Classification

Visualizing Data for Classification

In this lab, your goal is to explore a dataset that includes information about German bank credit to understand the relationships for a **classification** problem. In classification problems the label is a categorical variable.

Visualization for classification problems shares much in common with visualization for regression problems. Colinear features should be identified so they can be eliminated or otherwise dealt with. However, for classification problems you are looking for features that help **separate the label categories**. Separation is achieved when there are distinctive feature values for each label category. Good separation results in low classification error rate.

By the completion of this lab, you will:

1. Examine the imbalance in the label cases using a frequency table.
2. Find numeric or categorical features that separate the cases using visualization..

Lab Steps

1. Make sure that you have completed the setup requirements as described in the Lab Overview section.
2. Now, run jupyter notebook and open the "VisualizingDataForClassification.ipynb" notebook under Module 2 folder.
3. Examine the notebook and answer the questions along the way.

Question 1

1.0/1.0 point (graded)

From the created plots, which two features seem to separate the good and bad credits?

- [x] payment_pcnt_income
- [ ] number_loans
- [ ] age
- [x] loan_amount

✔

Submit You have used 2 of 2 attempts

Question 2

1.0/1.0 point (graded)

From the created plots, which feature seems to separate the good and bad credits?

- ○ foreign_worker
- ○ telephone
- ○ job_category
- ◉ checking_account_status ✔

Submit You have used 2 of 2 attempts

Learn About Verified Certificates