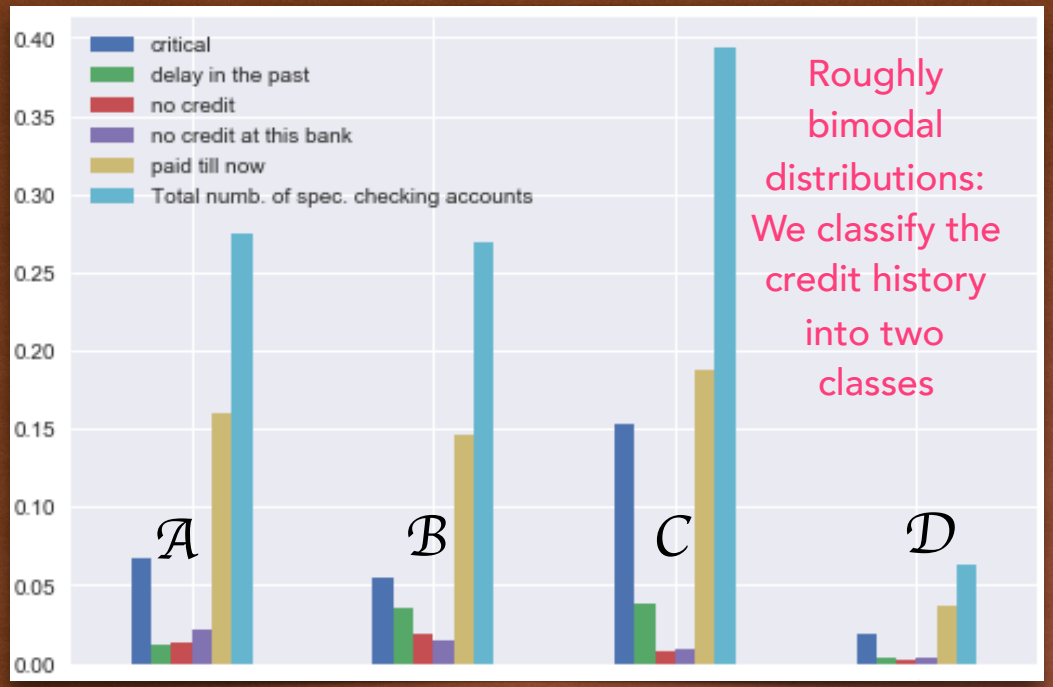
PREDICTION FOR DEFAULT ON LOAN PAYMENT

Number of customers (per five thousand)



Customers with balance > $10 K USD have the maximum instances of both, critical delay and on-time payments

Our simplified default on payment prediction model (Three features):

Credit history: good (0) and bad (1)

Job status: employed/ employable (0) and unemployed/ unemployable (1)

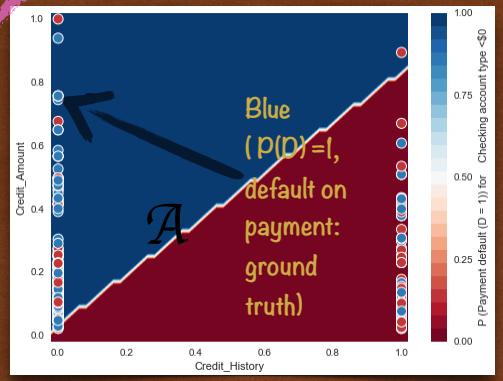
Credit amount (normalised between 0 and 1)

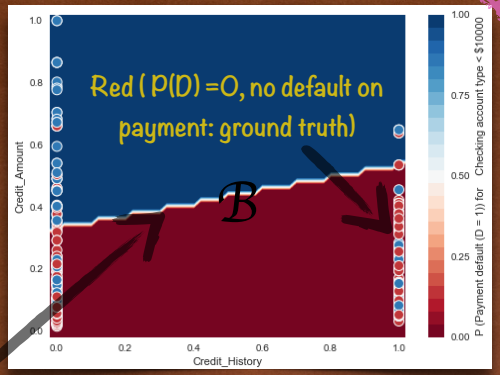
Probability for default on loan payment = P(D)
Scenario 1: P(D) ~ f(Credit history, Credit Amount)
Scenario 2: P(D) ~ f(Credit Amount,Job Status)

Its a classification problem: We use Logistic Regression to train our model in the above two scenarios to predict if a customer will have default on payment

Probability for default on loan payment = F(L)

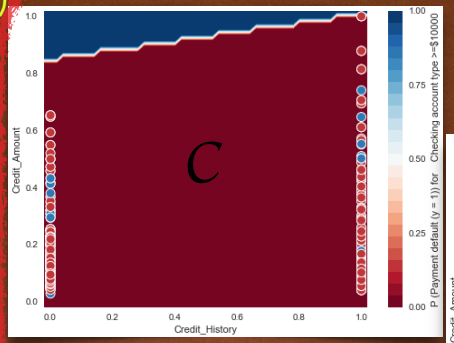
Scenario 1: P(D) ~ f(Credit history, Credit Amount)

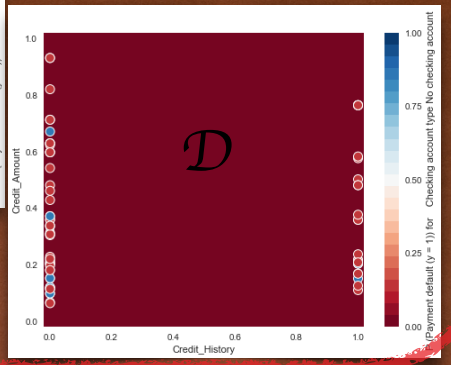




Predicted Classification boundary Accuracy ~ 60 %

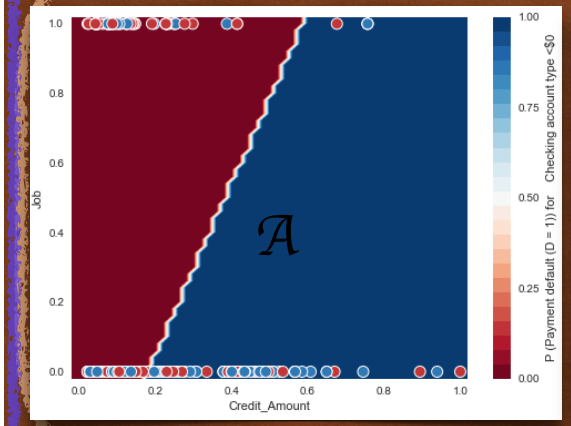
Accuracy ~ 85 %

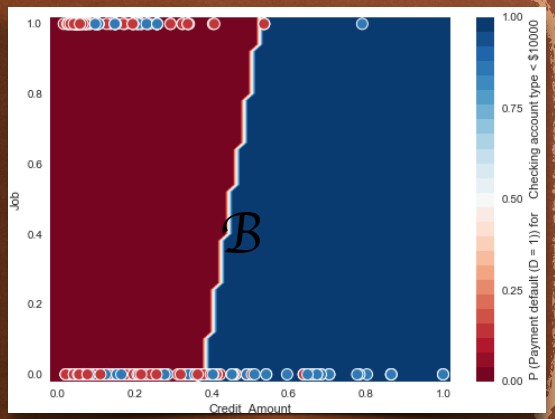




Probability for default on loan payment = F(L)

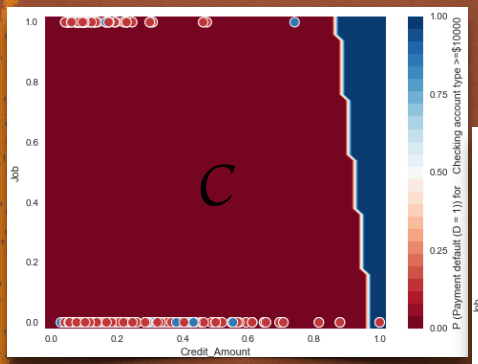
Scenario 2: P(D) ~ f(Credit Amount,Job Status)

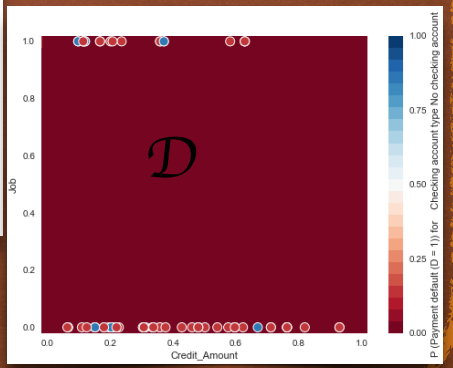




Accuracy ~ 60 %

Accuracy ~ 85 %





Summary: Scenarios 1 and 2:

Our simple predictive model can predict within certain accuracy the probability of customer to have default on payment. The accuracy of our simple model is better in instances C and $\mathcal{D}$ because in those instances the credit history does (roughly) split into two peaks i.e. good or bad. For cases $\mathcal{A}$ and $\mathcal{B}$ we have to train our model with more features to further increase its accuracy.