HW 7b

Andres Potapczynski (ap3635)

Plan for final project

For the final project we plan to analyze the risk in all the mortage contracts that a startup in Latin America has signed. For each contract, they have shared with us multiple covariates: age, interest rate, maritial status, months employed, credit score, contract length, geolocation, an many more (over 30). We want to understand how these covariates interact and predict the probability of default (over three months of no payment).

Bayesian model

Thus far, we are planning on doing a Generalized linear model as the ones shown in chapter 16. A logistic regression appears suitable for our current task but we are also thinking of exploring basis functions alternatives (like splines) or gaussian process models. All this only if suitable.

Posterior Predictive Checks

The posterior predictive checks that we have discussed so far relate to evaluating which types of mortages our simulations predict as the more likely to default. For example, mortages from people that are unemployed, married and have a low credit score would be expected to have the highest probability of default. Our posterior predictive check will compare the sd predicted in our model against the sd seen in the data. Another posterior predictive check would be to see the same risk ranking of cities in our data base

Model Evaluation

There are two main alleys of how we are going to test our models: (1) CV and (2) Sensitivity analysis. Since our data base comprises 70K mortgage contract, we can bare the computational expense of performing CV. Also, we are going to try different model specifications and also different prior distributions.

Data Collection

We asked the startup to give us most of the information that they record about their customers. Our understanding is that this guarantees ignorability. We might be needing to ask for, perhaps, historical information if we find that we need so.