Abstract: Divorce Analytics

Team: Akhilesh Soni, Zach Zhou

Problem Statement: Given a model that can predict the likelihood of divorce and explain contributing factors, marriage counseling therapists can have a better understanding of the direction in which a marriage is headed, and offer counseling accordingly. In this work, we apply machine learning techniques towards predicting whether or not a couple will divorce based on responses to a questionnaire called the Divorce Prediction Scale (DPS). We also identify the most telling questions on the survey to explain what the key factors are to explaining whether a couple might divorce.

Data: We investigate the Divorce Predictors dataset, which consists of the responses of 170 couples (84 divorced, 86 married) to the DPS. The DPS contains 54 questions with responses on a scale of 0 to 4, where 0 means "strongly disagree" and 4 means "strongly agree."

Methods: We compare the performance of logistic regression, KNN, decision trees (CART and optimal classification tree), random forest, and SVM on this dataset. We also developed a novel KNN variant based on Integer Programming, in which features are weighted, and the learning task is to find weights for each of the features in order to maximize in-sample accuracy.

Results: Our EDA shows a clear separation in the mean feature values; divorced couples were much more likely to agree with statements on the questionnaire. Hence, most of the models gave an out-of-sample accuracy of over 99%. To identify the most telling questions on the survey, we used ℓ_1-regularized logistic regression which gave 9 features with nonzero regression coefficients, implying those questions are the most important for the practitioner in determining the outcomes.

Conclusions: The data shows a clear separation between divorced and married couples feature values. A simple model (logistic regression) is sufficient to predict marriage outcome with high accuracy, identify important features, and explain their importance. Therefore, in practice, a marriage counselor would likely get the most use out of a logistic regression model.