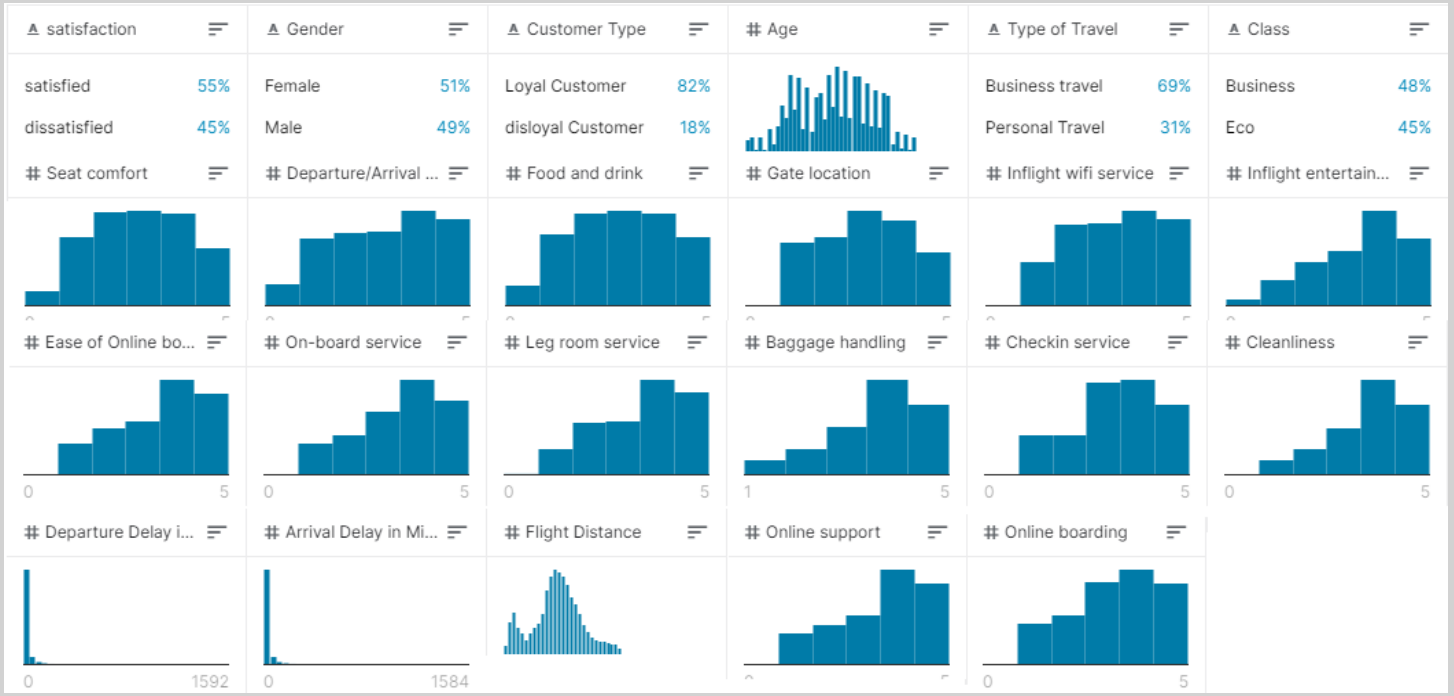
Passenger satisfaction on flights

Patryk Choma, 1617474
Lorenzo Chicca, 1708956
Nicolò D'Evangelista, 1698229
Rigels Hysaj, 1706263

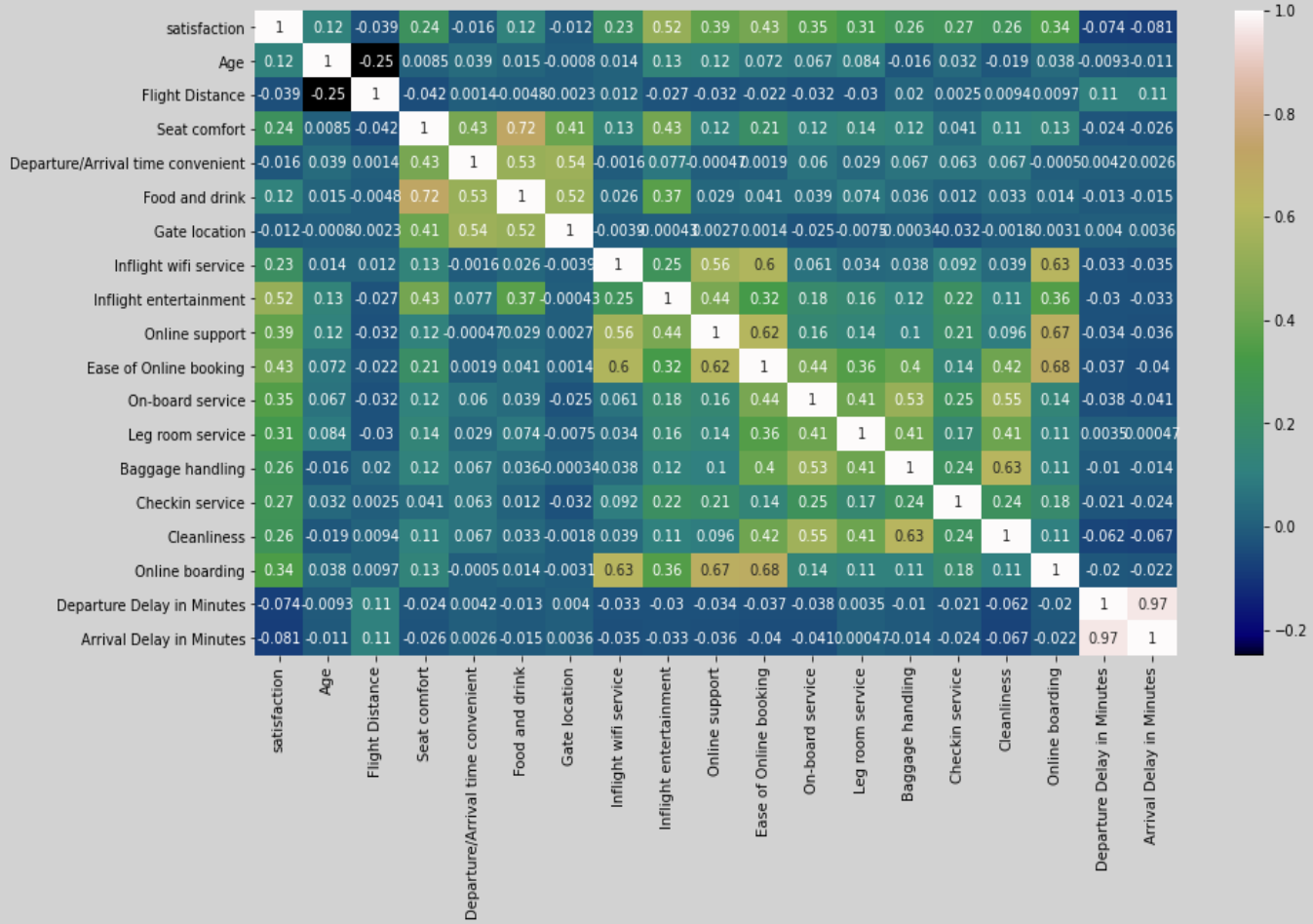
Dataset

- Dataset: https://www.kaggle.com/sjleshrac/airlines-customer-satisfaction
- Feature rich, many from surveys



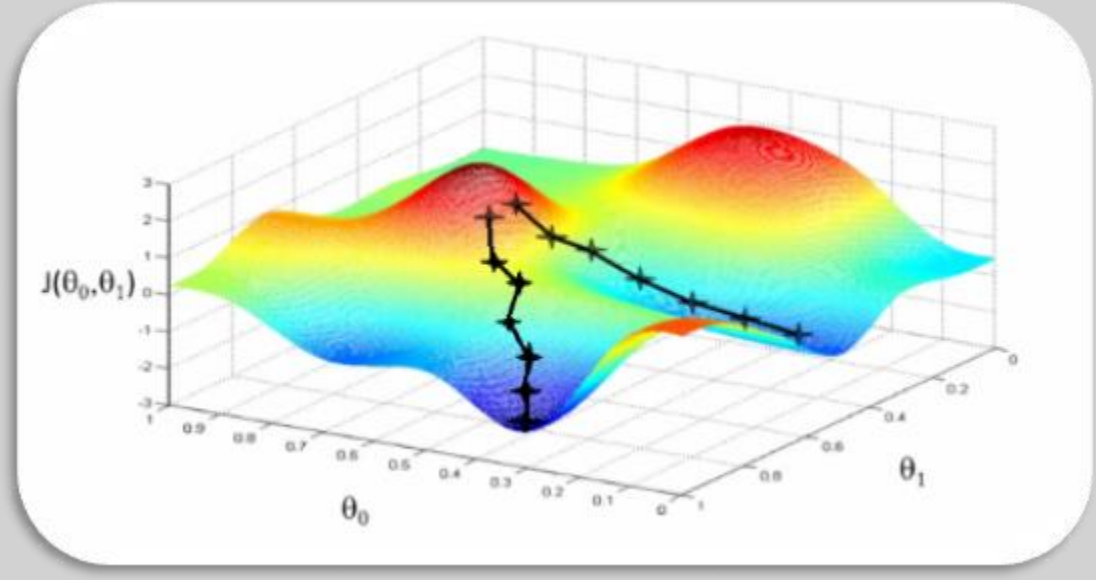
Preprocessing

- We get rid of features not correlated with "satisfaction"
- We delete the very few rows with missing values
- Rescaled all features into [0, 1] range
- One-Hot encoded the few qualitative features



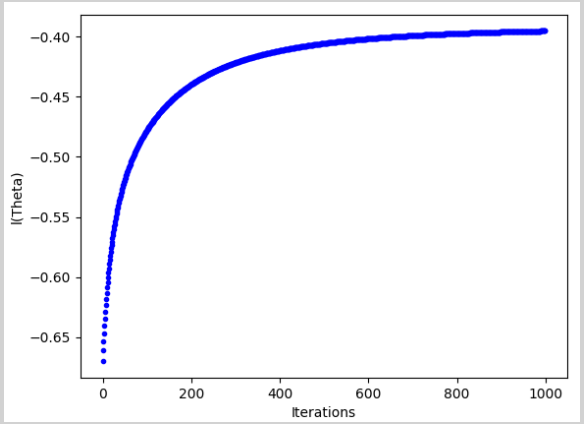
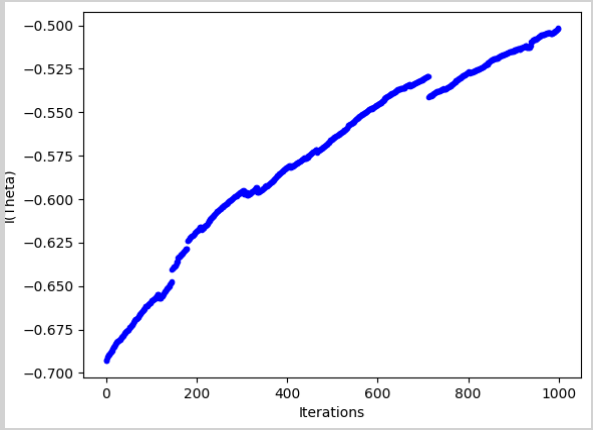
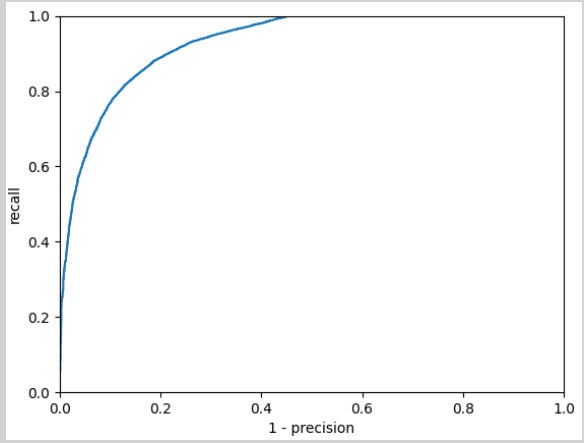
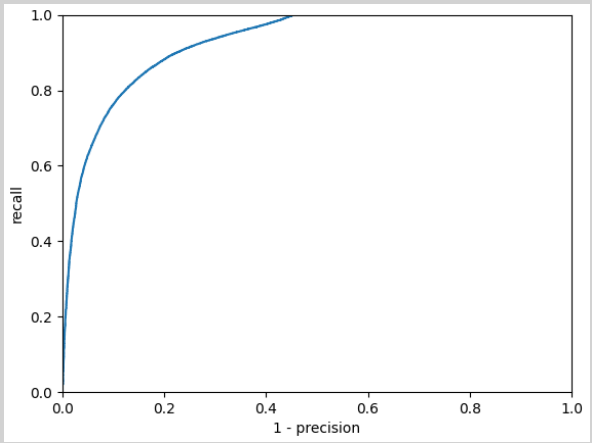
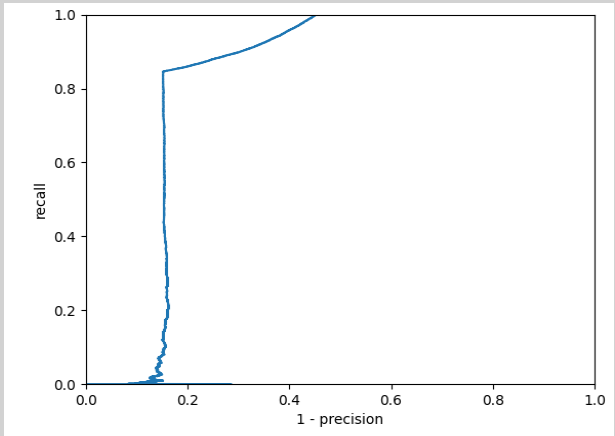
Model & Training

- Train/Test split is 70%/30%
- We used 3 different models
  - Gradient Ascent
  - Newton Method
  - Gaussian Discriminant Analysis



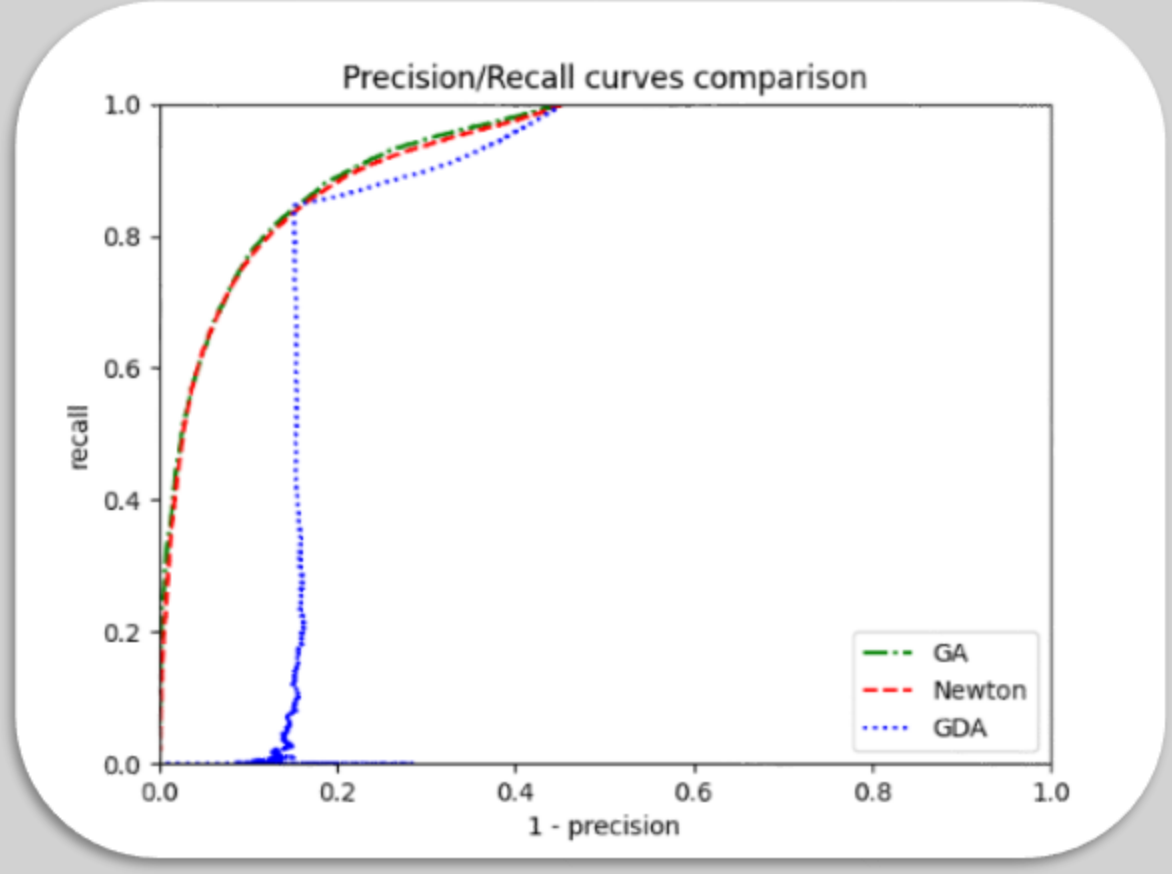
Results

- Convergence over iterations
- Precision / Recall curves

Gradient Ascent	Newton Method	Gaussian Discriminant Analysis
		
		

Conclusions

- GA is best overall
- Newton has good results only with a lucky initialization
- GDA has the highest peak in accuracy
- Overall all methods achieve > 80% accuracy



Literature and related projects

- CS229 Lecture notes - Andrew Ng
- Identifying and attacking the saddle point problem in high-dimensional non-convex optimization
- Bimarsha Khanal - Notebook about the same dataset on Kaggle