

Boosters, Digital Reputation

Соловьев, ММП, 2019

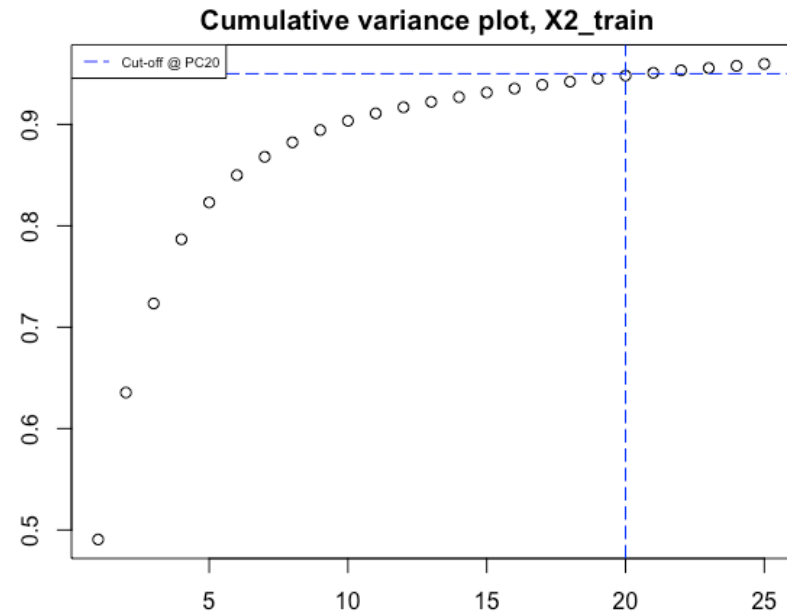
Резюме

Результат 0,576

- Наборы данных:
 - X1
 - X2
 - X1+X2+X3
- Преобразования данных:
 - PCA
 - Нормализация на $[0,1]$
 - Использование базовых статистик X2 и X3
- Модель итоговая:
 - Бустинг
- Низкий показатель AUC :
 - Автоэнкодер
 - Деревья
 - Стекинг бустинга

Данные

- Не нашел связей между переменными в X1
- Не нашел связей между X1 и X2 (за исключением id)
- Не нашел связей между X2 и X3
- X3 – характер данных не нашел, использовал базовые статистики – средняя, медиана, стандартное отклонение.



- X2 сворачивается PCA

Наборы данных:

- X1
- X2 – PCA + статистики X2
- X1 (наиболее значимые) + X2 (наиболее значимые) + статистики X3

Модели

- Пробовал autoencoder, GBM, RF, регрессию
- Выигрывает Бустинг
- На простейшем блендинге получил ошибку переобучения

