digikala GROUP

Digikala AI & Big Data Team

digikala

AI in Digikala Problems

- There are multiple types of AI usage at Digikala:
  - Better business insight
  - AI as a service for users
  - AI helps to run the business processes automatically

AI for Better Business Insight

- Demand prediction
- Next business step prescription
- Trending products detection
- Best Warehouse Operation Algorithim

AI As a Service for Users

- Search
- Recommendation
- Search Autocorrect
- Cons and Pros of Products

Running the Business Processes Automatically

- Find rules and hard code them.
- Use AI to extract the patterns and create a model.

Running the Business Processes Automatically(2)

- Auto Comment Moderation
- Background Removal
- Auto Price Moderation
- Bin Packing

Demand Prediction Problem

- Next 30 days daily sales count of products
- 60days daily sales count of 5000 products
- Prediction for warehouses and promotions

ML Problem Solving Pipline

1. Dealing with noises (logicaly wrong data)

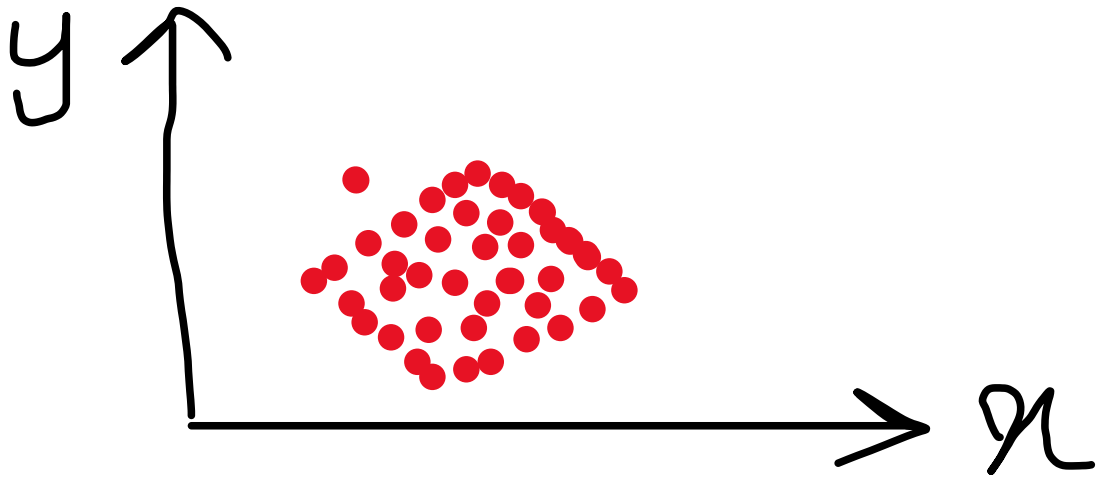
2. Dealing with missing values

    - Remove high null rated features or records
    - Fill with mean or median
    - Fill with mode
    - Use imputation techniques

3. Categorical variables to numerical

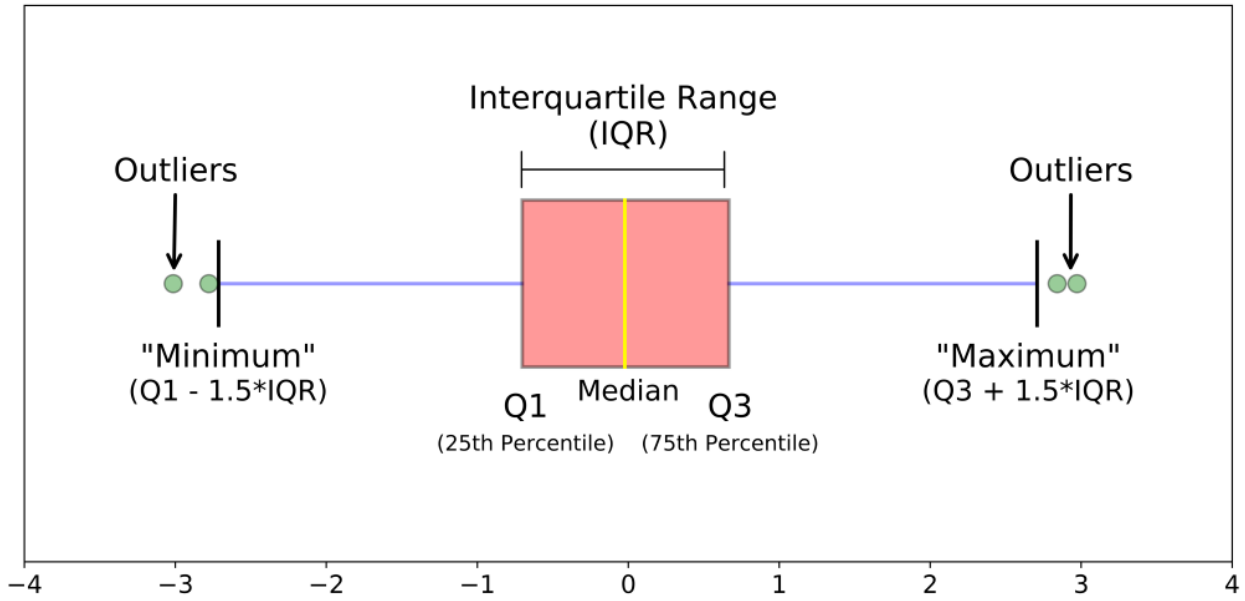
ML Problem Solving Pipline

4. Identifying abnormal events in the data



ML Problem Solving Pipline

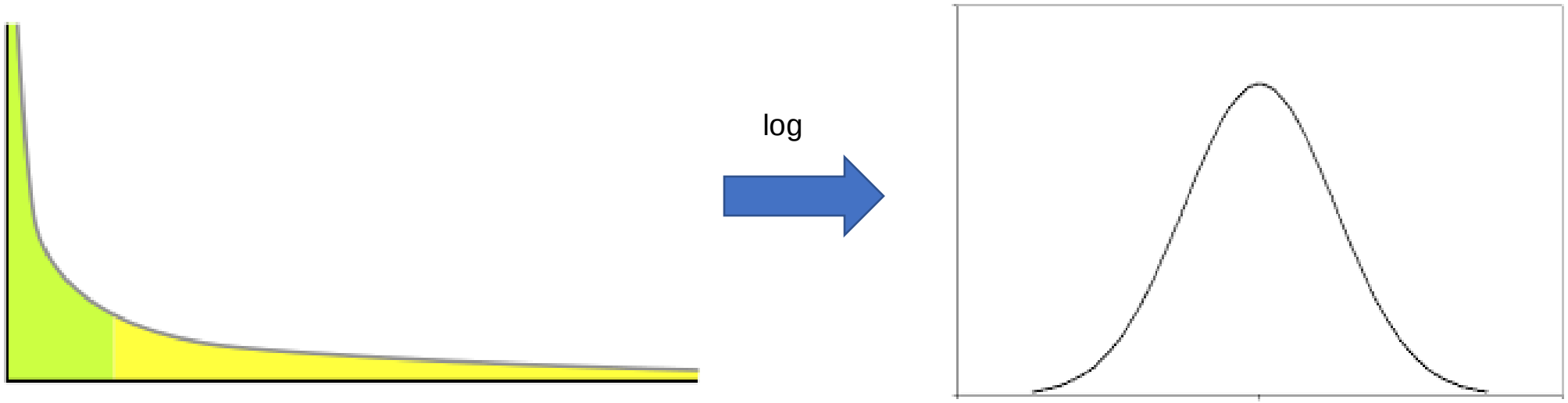
5. Features normalization
    1. min−max
    2. z-score
    3. You can use the upper and lower whiskers of the boxplot to prevent the influence of outliers.



ML Problem Solving Pipline

5. Features normalization

- Distribution transformation:
  - Equalization



ML Problem Solving Pipline

5. Features selection

   1- Use PCA or correlation matrix to remove redundant features

   2- Use mutual information as supervised way (random forest)

   3- Use forward or backward elimination

   4- NN supervised feature extraction

ML Problem Solving Pipline

6. Check balancing of data

   1- Use bootstrapping for up sampling

   2- If data cannot be up sampled use Decision Trees

   3- In Neural Networks use weights for lost functions or up sample in the batches

ML Problem Solving Pipline

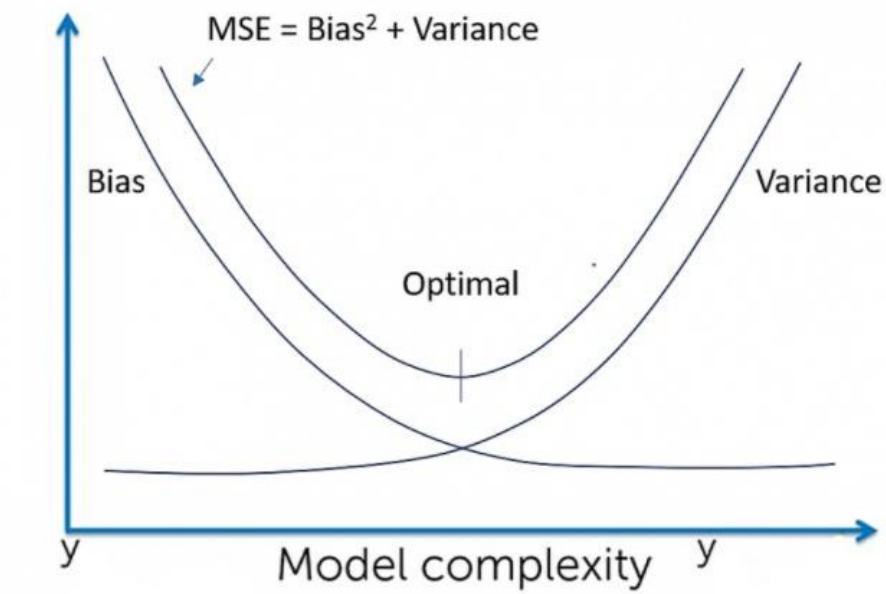
7. Building and comparing several different ML models
   - KNN (with different k)
   - Random Forest
   - XGBoost
   - GLM regression
   - Naïve Bayes
   - Bayes
   - SVM
   - NN

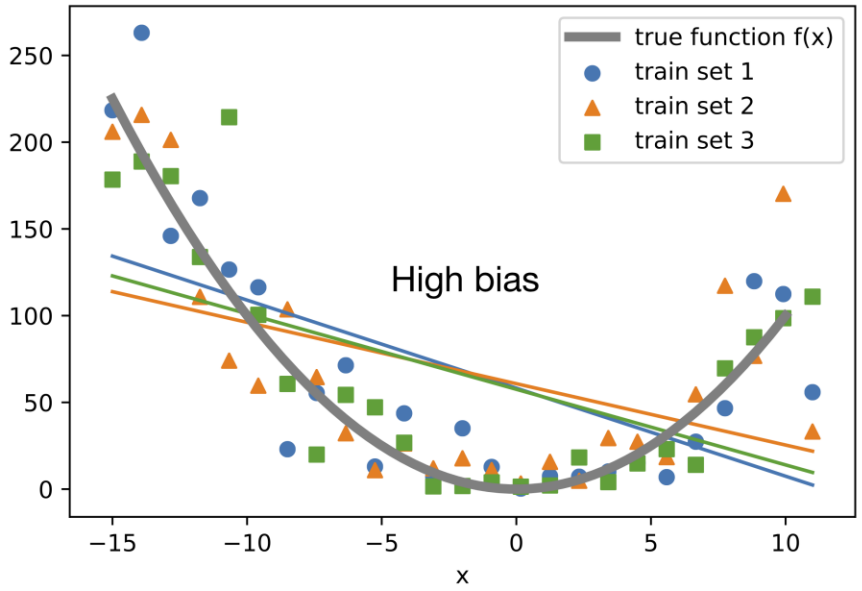
   * Use cross validation to overcome the selection bias

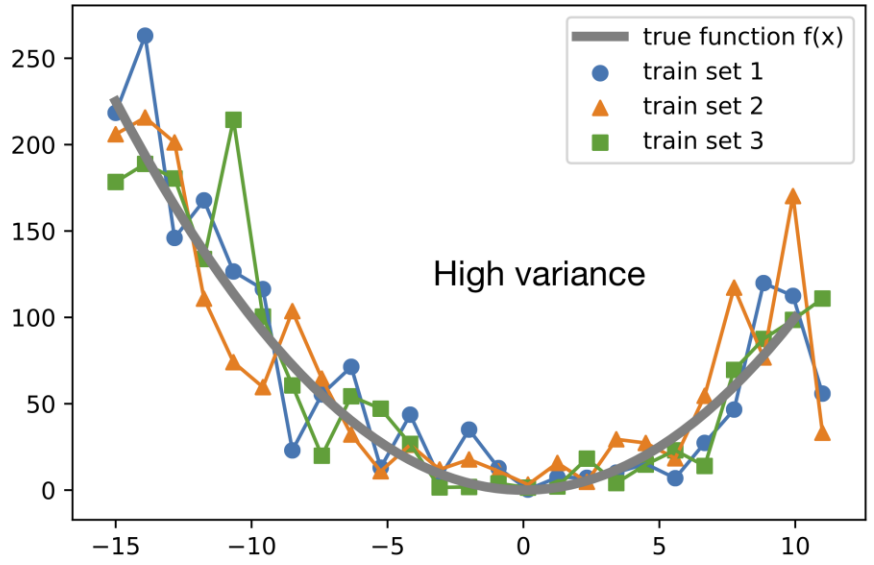
ML Problem Solving Pipline

8. Analysis of models and their results

- Check biase
- Check variance

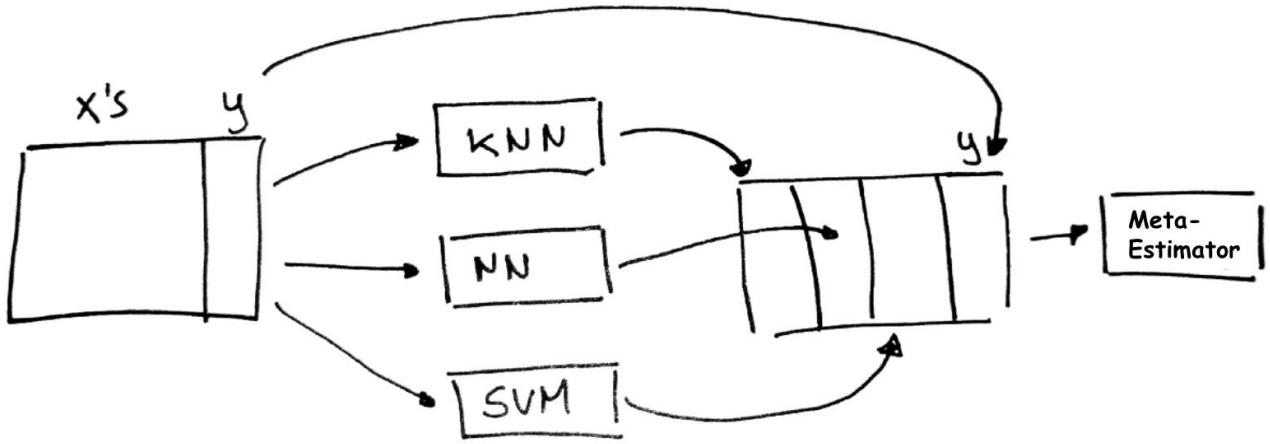






ML Problem Solving Pipline

9. Don't forget ensemble multiple models



.10 Use multiple metrics to see difference aspects.
Regression: MAPE, MSE, MAE
Classification: ACC, Recall, Precision

* Plot residuals vs each feature.

Questions?

https://careers.digikala.com/

digikalalife@digikala.com

@Digikala.life

digikala