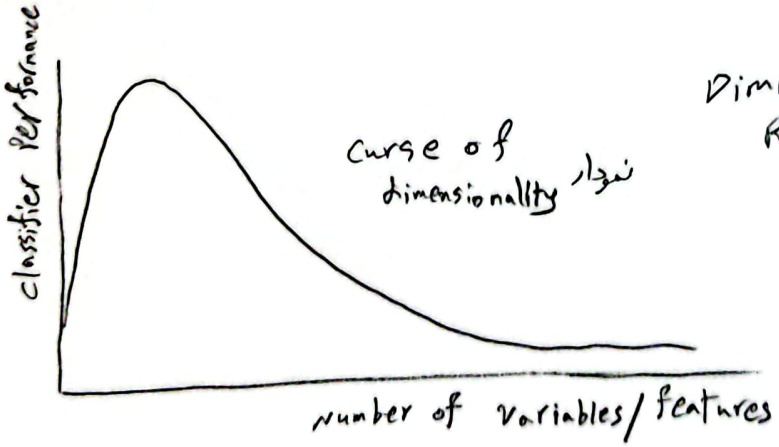
>> Binning sparse values

→ grouping or categorizing continuos or sparse numerical data into discrete bins or intervals.

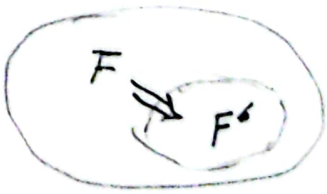
- Identify the sparse values
- Define the bins
- Assign values to Bins

Binning انواع:
- Equal-width Binning
- Equal-frequancy Binning
- Custom Binning
- Quantile Binning (P. quartiles, percentiles)

>> Feature selection



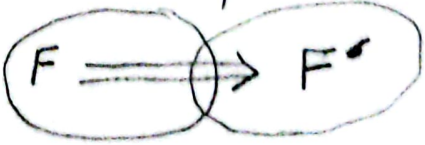
Dimensionality Reduction : راه‌حل



(★) Feature selection:

$$\{f_1, \dots, f_i, \dots, f_n\} \xrightarrow{\text{f.selection}} \{f_{i_1}, \dots, f_{i_j}, \dots, f_{i_m}\} \;\Big|\; i_j \in \{1, \dots, n\};\ j = 1, \dots, m \quad i_a = i_b \Rightarrow a = b;\ a, b \in \{1, \dots, m\}$$

(★) Feature Extraction/Creation



$$\{f_1, \dots, f_i, \dots, f_n\} \xrightarrow{\text{f.extraction}} \{g_1(f_1, \dots, f_n), \dots, g_i(f_1, \dots, f_n), \dots, g_m(f_1, \dots, f_n)\}$$

>> Denoising

انواع
- image Denoising
- signal Denoising → remove noise from time-series data
- Data Denoising → { outlier detection, correction

تکنیک ها
- Filtering
  - mean / Box Filter
  - Guassian Filter
  - median Filter
- wavelet Transform → DWT (Discrete Wavelet Transform)
- Fourier Transform
- Non-Local means (NLM)
- Auto encoders

>> Sampling

اهداف اصلی sampling
- کارایی و بهصرفه بودن از لحاظ هزینه و زمان
- امکان تایید بودن بررسی کل جامعه آماری
- دقت و precision ← بستگی به روش ما دارد

طبقه بندی کلی روش های نمونه برداری
- probability sampling
  - simple random sampling (SRS)
  - Systematic sampling
  - stratified sampling
  - cluster sampling
  - Multistage sampling
- Non-probability
  - convenience sampling
  - Judgemental or Purposive sampling
  - Quota sampling
  - snowball sampling

Simple Random Sampling (SRS) → هر عضو احتمال انتخاب شدن یکسان دارد

systematic sampling → انتخاب هر k امین عضو

stratified sampling → Dividing a population by age groups & randomly sampling from each age

cluster sampling → انتخاب یک مدرسه از چندین مدرسه و سپس پرسش از تمام دانش آموزان آن مدرسه

multistage sampling → selecting states, then cities within those states, and the individuals within those cities.

convenience sampling → نمونه‌ها براساس راحتی دسترسی انتخاب می‌شوند.

Judgmental sampling → selecting experts in a field for an opinion survey

Quota sampling → اطمینان حاصل کردن از اینکه نمونه‌ها شامل مقدار مشخصی از هر گروه باشد

snowball sampling → Surveying a hidden or hard-to-reach population, such as drug users.